농학박사학위논문

The use of next generation sequencing to decipher metagenome of domestic animals

차세대 시퀀싱을 이용한 가축들의
메타제놈에 대한 연구

2015년 2월

서울대학교 대학원
농생명공학부 동물생명공학전공
임 수 연

The use of next generation sequencing to decipher metagenome of domestic animals

By

Sooyeon Lim

Supervisor: Professor Heebal Kim

February, 2015

Department of Agricultural Biotechnology

Seoul National University

차세대 시퀀싱을 이용한 가축의 메타제놈에 대한 연구

지도교수 김 희 발

이 논문을 농학박사 학위논문으로 제출함

2015 년 2 월

서울대학교 대학원

농생명공학부 동물생명공학전공

임 수 연

임수연의 농학박사 학위논문을 인준함

2015 년 2 월

위 원 장 이 창 규 (인)

부위원장 김 희 발 (인)

위 원 김 경 모 (인)

위 원 윤 숙 희 (인)

위 원 조 서 애 (인)

ABSTRACT

The use of next generation sequencing to decipher metagenome of domestic animals

Sooyeon Lim

Department of Agricultural Biotechnology

The Graduate School

Seoul National University

This study investigated metagenomic characteristics of microbial communities in the domesticated poultry through the combination of next generation sequencing (NGS) and bioinformatic pipelines. Among various kinds of commercial animals, this study placed its focus on the bovine rumen, chicken feces and goat rumen fluid, respectively. In more detail, structure of microbial population in the various samples of rumen (bovine/goat) and feces (chicken) was analyzed and then, its phylogeny and specific gene such as cellulase was identified. Sequencing the genomes of rumen microbes, determining the role of the genes and identifying its potential applications are the great deal for researchers to understand the microbiology of the rumen. Chicken feces have generally been known to be closely associated with contamination of poultry and health safety. Goats have unique habits, which

include feeding on unconventional tree leaves. Thus, goats are expected to host distinct bacterial communities with cellulose-degrading enzyme activity in their rumen.

In cahpter1, basic background and necessity were reviewed the series of worked in this doctoral dissertation, which showed its pivotal role of microbiomes in the commercial animal in various respects. Furthermore, metagenomics and genetics/genomics can provide a significant clue to these microbial population.

In chapter 2, the microbial community structure of rumen solid and rumen liquid of cattle rumen was analyzed using high-throughput pyrosequencing of 16S rRNA gene PCR amplicons and a subsequent bioinformatics pipeline. A 16S rRNA gene clone library identified abundant communities belonging to specific bacterial groups in the rumen. The diversity results suggested that the specific bacterial groups was found in both samples with a slight difference. Bacteroidetes, Firmicutes and Fibrobacteria were present in higher concentrations in rumen solid than in rumen liquid, indicating their major role in the degradation of plant fiber. Other groups identified include Proteobacteria, which are responsible for playing a greater role in rumen metabolism; *Mollicutes* class of Firmicutes, which metabolizes imported sugars to short chain fatty acids; and *Prevotellacea*, which are crucial for the breakdown of proteins. These biological function of identified microbial groups mentioned above have also coincided with other previous studies.

In chapter 3, fecal microbial community of chicken was quantitatively analyzed using next generation sequencing (NGS) techniques and bioinformatic analyses with metagenomic tools such as MOTHUR, MEGA6 etc. during a relatively short growth time of 35 days. The diversity of microbial community at the genus level increased during the five week growth period (from 30 to 87 identified genera). Despite the diversity, only a few dominant bacteria groups (over 80%) were identified in each fecal sample, which were completely different from each other. These results suggested that chicken fecal microbiome is a dynamic system with a differentiated population structure under a restricted number of higher taxa.

In chapter 4, protein domains with cellulase activity in goat rumen microbes were investigated using Illumina sequencing and bioinformatic analyses with metagenomic tools such as METAIDBA, HMMER and Interproscan etc. After the complete genome of the goat rumen microbe was obtained using a shotgun sequencing method, 217,892,109 pair reads were filtered using METAIDBA. These filtered contigs were assembled and annotated using blastN against the NCBI nucleotide database. As a result, a microbial community structure with 439 genera was analyzed, among which *Prevotella* and *Butyrivibrio* were the dominant groups. In parallel, 201 sequences related with cellulase activities (EC.3.2.1.4) were obtained through blast searches using the enzyme.dat file provided by the NCBI database. 28 protein domains with cellulase activity were identified using the HMMER package. Cellulase activity protein domain profiling showed that the major

protein domains such as lipase GDSL, celluase, and Glyco hydro 10 were present in bacteria with strong cellulase activities. Furthermore, correlation plot clearly displayed the strong positive correlation between some protein domain groups, which was indicative of microbial adaption in the goat rumen based on feeding habits. Recent studies clearly reported that intestinal microbome was closely correlated with the traits of host such as obesity and growth. Therefore, it can be easily expected that analysis of intestinal microbial structure in commercial animals can provide an insight to the livestock industry in terms of fiber digestion and growth.

Keywords: Next generation sequencing (NGS), Metagenome, Ruminant, Microbial community

Student number: 2010-30316

Contents

List of Tables

List of Figures

List of Appendix

Chapter 1. Literature Review

1.1 Sequencing Techniques

DNA sequencing is the process of determining the precise order of nucleotides in a DNA strand. It includes any method or technology which is used to determine the order of the four nucleotide bases such as adenine (A), guanine (G), cytosine (C), and thymine (T). The advent of rapid DNA sequencing methods has greatly accelerated a revolution in the field of molecular biology.

1.1.1. First Generation Sequencing

In 1977, the first two widely-known DNA sequencing methods were reported by Maxam and Gilbert(Maxam, 1977) and Sanger and collaborators(Sanger et al., 1977) , respectively. Despite of each basic principle of chemical cleavage (Maxam, 1977) and chain terminator sequencing (Sanger et al., 1977), both approaches rely on the separation of the mixture of DNA fragments of various sizes on polyacrylamide (PAA) slab gels. The Sanger DNA sequencing technique is based on DNA synthesis with incorporation of normal dNTPs as well as ddNTPs causing a termination of the newly synthesized DNA molecule (Figure 1.1). Procedure automation of Sanger sequencing was motivated for the purpose of fast reading the massive DNA sequences. For example, Applied Biosystems, Inc. (ABI) made two major improvements of DNA labelling with fluorescent dyes and computer based data analysis. ABI 370 instrument, containing PAA gel with 16 lanes, adopted

fluorescent dyes for the labelling of DNA primers instead of radioactive isotopes. The fluorescence could be detected in real time through laser excitation close to the bottom of the gel. In both one and four dyes configuration, four electropherograms or chromatograms could be obtained. The sequence of the analyzed fragment could be produced by superimposing. Over the next two decades, ABI sequencers were significantly improved (Table 1.1). The number of lanes in PAA gel-based models increased from 16 (ABI 370A) to 96 (ABI 377). The first capillary sequencer was model ABI 310, with one capillary, and the model 3730xl had 96 capillaries. At the same time, the length of the reads increased from 350 (ABI 370A) to over 900 (ABI 3730xl), while the run times decreased from 18h to 3h. The invention of model 3730 increased the speed of the sequencing of the human genome. The major drawbacks are the price per base and the problems related to cloning and sequencing of regions containing repetitive sequences(Gužvić, 2013).

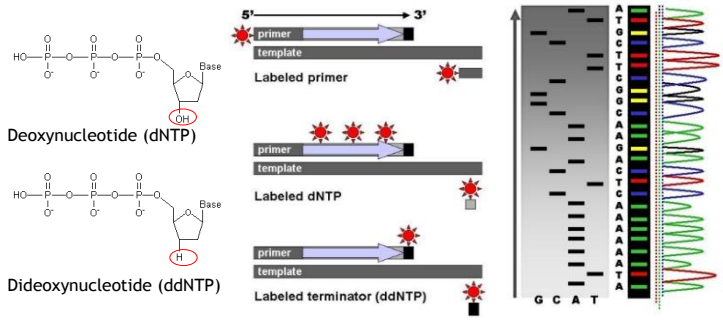


Figure 1.1. Principles of 1st generation DNA sequencing. DNA fragments are labelled with a radioactive or fluorescent tag on the primer, in the new DNA strand with a labeled dNTP, or ddNTP. Technical variations of chain-termination sequencing include tagging with nucleotides containing radioactive phosphorus for radiolabelling, or using a primer labeled at the 5' end with a fluorescent dye (Source : http://en.wikipedia.org/wiki/Sanger_sequ encing).

Table 1.1. Comparison of DNA sequencers

Techniques	Classification
Maxam-gilbert	1st generation
Sanger	1st generation
First Automated DNA Sequencer ABI370 (373)	1st generation
Pharmacia ALF1	1st generation
ABI377 (Up to 96 lanes)	1st generation
First Capillary DNA Sequencer ABI 310	1st generation
First 96 Capillary instruments MegaBace ABI 3700	1st generation
ABI 3100, 16 Capillary	1st generation
ABI 3730, 48 or 96 Capillary	1st generation
Genome Sequencer GS20, 454 Life Science, Roche	NGS
Genome Analyzer, Solexa/Illumina	NGS
SOLiD (Applied Biosystems)	NGS
Helicos (Helicos Bioscience)	NGS
PacBio (PacBio Bioscience)	NGS
Nanopore (Oxford Nanopore)	NGS

1.1.2 Next Generation Sequencing (NGS)

The automated Sanger method described in the previous section is considered as a 'first-generation' technology. Next generation sequencing (NGS) generally refers to non-Sanger based high-throughput DNA sequencing technologies. The key feature of NGS is the production of large volumes of sequence data, which provides a primary advantage over conventional methods in terms of throughput, scalability, speed and resolution. NGS implements cyclic-array sequencing whose concept can be summarized as the sequencing of a dense array of DNA features by repeating cycles of enzymatic manipulation and data collection based on imaging. This second-generation technology has recently been realized in commercial product such as 454 sequencing (Roche Applied Science), Solexa technology (Illumina), SOLiD platform (Applied Biosystems), Helicons (Helicos Bioscience), PacBio (PacBio Bioscience) and Nanopore (Oxford Nanopore). Although these platforms are quite diverse in sequencing biochemistry as well as in how the array is generated, their work flows are conceptually similar (Figure 1.2). Preparation of library is produced by random fragmentation of DNA, followed by in vitro ligation of general adaptor sequences. The generation of clonally clustered amplicons to serve as sequencing features can be accomplished by a number of approaches, including in situ polonies, emulsion PCR or bridge PCR.

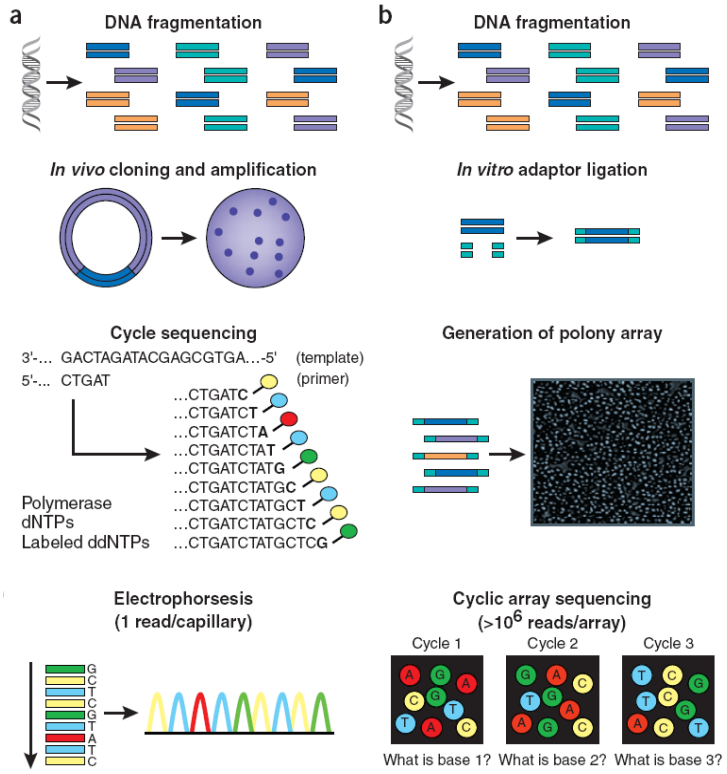


Figure 1.2. Work flow of (a) conventional high-throughput Sanger sequencing and (b) second-generation sequencing (Shendure & Ji, 2008).

1.1.2.1. Pyrosequencing and 454 FLX system

Pyrosequencing is a method DNA sequencing based on the sequencing by synthesis principle, which is outlined in Figure 1.3. In the pyrosequencing system included 4 enzymes such as the Klenow fragment of DNA polymerase I, ATP sulfurylase, Luciferase and Apyrase. Combination of reaction also includes the sequencing template with an annealed primer for starting material of the DNA polymerase, enzyme substrates adenosine phosphosulfate (APS) and d-luciferin. At a time, 4 nucleotides are added repeatedly and camera detects the light generated (Ahmadian et al., 2006).

The enzymatic reactions utilized in the pyrosequencing technology as followed. At the first, DNA polymerization, incorporating into the growing DNA strand.

$$(DNA)_n + dNTP \rightarrow (DNA)_{n+1} + PP_i$$ The

inorganic pyrophosphate, PPi, discharged by the Klenow DNA polymerase serves as a substrate for ATP Sulfurylase that causes ATP:

$$PP_i + APS \rightarrow ATP + SO_4^{2-}$$

Through the third and fourth reactions, the ATP is produced the light by Luciferase. Therefore if the proper nucleotide is added to the reaction compound, light is generated.

$$\text{Luucifersae} + \text{D-luciferin} + \text{ATP} \rightarrow \text{Lucfierase-luciferin-AMP} + PP_i$$

$$\text{Lucfierase - luciferin - AMP} + O_2$$
$$\rightarrow \text{Luciferase} + \text{oxyluceferin} + \text{AMP} + CO_2 + \text{Light}$$

Apyrase removes unused nucleotides and ATP before the additions of different bases insert.

$$\text{ATP} \rightarrow \text{AMP} + 2P_i \ , \ \text{dNTP} \rightarrow \text{dNMP} + 2P_i$$

This degradation between base additions is crucial for synchronized DNA synthesis asserting that the light signal was detected when adding a certain nucleotide only arises from incorporation of that specific nucleotide. DNA sequencer adopting this pyrosequencing by synthesis was firstly developed by 454 Life Sciences, which was later acquired by Roche. The main characteristics of 454 technology were pyrosequencing, miniaturized and massively parallelized using PicoTiter Plates. In this approach, libraries may be built by any method that gives rise to a combination of short, adaptor-flanked fragments. Clonal sequencing features are produced by emulsion PCR, with amplicons hold to the surface of 28-μm beads. After breaking the emulsion, beads are deal with denaturant to eliminate untethered strands, and then subjected to a hybridization-based abundance for amplicon-bearing beads. A sequencing primer is hybridized to the universal adaptor at the suitable site position and orientation then instantly adjacent to the start of unknown sequence. Limitation of the 454 technology relates to homopolymers. Since there is no closure moiety avoiding several consecutive incorporations at a given cycle, the length of all homopolymers must be deducted from the signal intensity. This is effect to a more error rate than the distinction of incorporation versus non-incorporation.

Therefore, the major error type for the 454 platform is insertion-deletion, other than substitution. Compare to other next-generation platforms, the major advantage of the 454 platform is read-length. Presently the per-base cost of sequencing with the 454 platform is much better than that of other NGS platforms such as Solexa and SOLiD. Nevertheless, it may be the way of choice for certain applications where long read-lengths are critical such as metagenomics (Shendure & Ji, 2008).

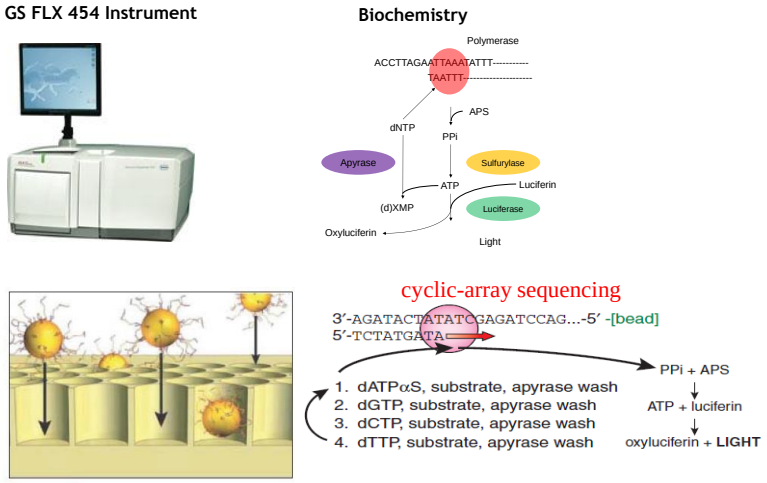


Figure 1.3. Schematic representation of the pyrosequencing and 454 GS FLX (Shendure & Ji, 2008).

1.1.2.2. Illumina Genome Analyzer

This sequencing method is commonly referred to as Solexa. Its main characteristic is that it is able to produce larger volume of data, compared to 454 sequencer. Libraries can be organized by any method that bring about a mixture of adaptor-flanked fragments, which are generally up to a number of hundred base-pairs. Through the bridge PCR, amplified sequencing features are generated. In this way, both forward and reverse PCR primers are binded to a solid substrate by a pliable linker, such that all amplicons generating from any single template molecule during the amplification stay immobilized and clustered to a single physical site on an array. On the Illumina platform, the bridge PCR is somewhat irregular in relying on alternating cycles of lengthen with *Bst* polymerase and denaturation with formamide. Each Clusters is consist of ~1,000 clonal amplicons from the result. Within each of eight lanes, millions of clusters can be amplified to distinguishable locations that are on a single flow-cell. During the same instrument run, eight independent libraries can be sequenced at once. After cluster occurrence, the amplicons are linearized and a sequencing primer is hybridized to a universal sequence flanking the target region. Consists of single-base extension with a modified DNA polymerase and a combination of four nucleotides in every cycles. (Figure 1.4). Chemically cleavable moiety at the 3’ hydroxyl site available only a single-base incorporation to occur in every cycle. In addition one of four fluorescent labels, which chemically cleavable are corresponds to the identity of every nucleotide. After single-base extension and acquisition of images in 4 channels, chemical

cleavage of these groups sets up for the next cycle. Read lengths are restricted such as unfinished cleavage of fluorescent labels or terminating moieties. The major error is substitution, other than insertions or deletions. Mean of raw error rates are on the order of 1–1.5%, but higher accuracy bases with error rates of less than 0.1% can be identified through quality metrics associated with each base-call (Shendure & Ji, 2008)

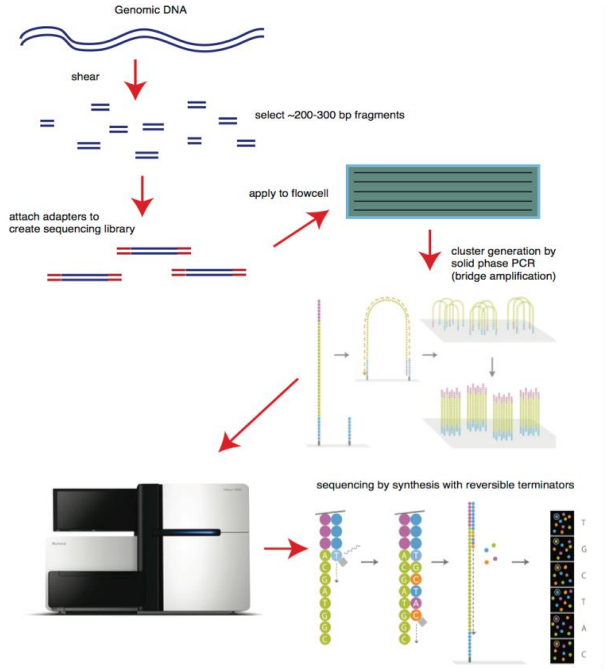


Figure 1.4. Schematic representation of the Illumina sequencing and HiSeq 2000 sequencer (Source: Stuart M. Brown, 2012 in Next Generation Sequencing).

1.1.2.3. SOLiD and SOLiDv4

DNA sequencer method of SOLiD *(Sequencing by Oligonucleotide Ligation and Detection)* is based on the sequencing by ligation (SBL) approach, in which DNA polymerase is replaced by DNA ligase. Libraries can be organized by any method that bring about a mixture of short, adaptor-flanked fragments. Through the emulsion PCR, Clonal sequencing features are produced with amplicons captured to the surface of 1-μM paramagnetic beads. Amplification results that beard beads through breaking the emulsion, selectively improved and immobilized to a solid planar substrate to generate a dense and disordered array. Sequencing by synthesis is driven by a DNA ligase. A universal primer is hybridized to the array of amplicon-bearing beads. Every cycle of sequencing includes the ligation of a degenerate population of fluorescent octamers (Figure 1.5). The octamer mixture is structured, in that the character of specific position with the identity of the fluorescently labeled octamers. After ligation, images are acquired each and collected data for the same base positions across all template-bearing beads. After removing the fluorescent label, octamer ligation enable sequencing of every 5th base. Completing these cycles the extended primer is denatured to reset the system for different set of positions. This platform involves the use of two-base encoding to error-correction scheme, rather than a single base. Each base position is queried twice and miscalls can be more readily identified. 454, SOLiD and the Polonator, are possible that sequencing on a high-density array

of very small beads may stand for the most straightforward opportunity to achieve tremendously high data densities (Shendure & Ji, 2008).

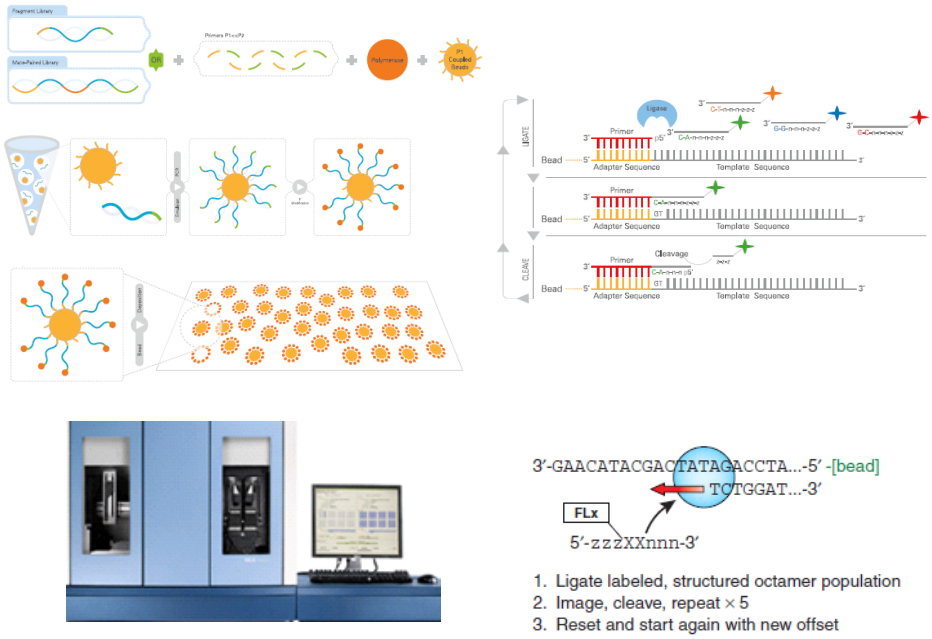


Figure 1.5. Schematic representation of the SOLiD sequencing and SOLiDv4 sequencer (Source: http://www.appliedbiosystems.com, Shendure & Ji, 2008).

1.1.2.4. Helicos single molecule fluorescent sequencing

The Helicos was the first commercial next generation sequencing platform which uses the principle of single molecule fluorescent sequencing (Harris et al., 2008). Briefly, the fragments of DNA strands are hybridized on disposable glass flow cells. Then, fluorescent nucleotides are added one after another with a terminating nucleotide which has the role of pausing the process until an image has been captured. From the captured image, single nucleotide from each DNA sequence can be determined. Finally, the fluorescent molecule is cut away, and the process is repeated until the sequencing of target fragments has been completed. In more detail, template libraries, which were prepared by random fragmentation and poly-A tailing, are captured by hybridization to surface-tethered poly-T oligomers and make a disordered array of primed single-molecule sequencing templates. At each reaction cycle, both DNA polymerase and a single species of fluorescently labeled nucleotide are added, which result in template-dependent extension of the primer-template duplexes immobilized on the surface. After images acquisition, chemical cleavage and release of the fluorescent label continues the subsequent cycle of extension and imaging. Harris and colleagues reported that several hundred cycles of this single-base extension make average read-lengths of 25 bp or greater (Harris et al., 2008). Notable aspects of Helicos sequencing are summarized as following. First, this sequencing is asynchronous like the 454 platform. That is to say some strands will fall ahead or behind others in a sequence-dependent manner. Chance also plays a part in this phenomenon, as some templates may simply

fail to incorporate on a given cycle despite of having the appropriate base at the next position. However, it can be said that dephasing is not an issue, and do not induce errors because these are single molecules. Second, there is no terminating moietyon the labeled nucleotides. Therefore, homopolymer runs can be an issue of importance. However, considering that single molecules are being sequenced in Helicos, this problem can be reduced by limiting the rate of incorporation reaction. Third, the accuracy of raw sequencing can be substantially improved by a two-pass strategy in which the array of single-molecule templates is sequenced and then fully copied. The newly synthesized strand is surface-tethered, and therefore, the original template can be removed by denaturing step. Sequencing, which was primed from the distal adaptor, generates the second sequence for the same template with the opposite orientation. At last, the dominant error type is reported to be deletion and the error rate was known to be 2~7% with one pass and 0.2~1% with two passes, respectively. However, error rates of substitution are reported to be substantially relatively lower in the range of 0.01~1% with one pass. With two passes, error rate of the per-base raw substitution approaches to 0.001%.

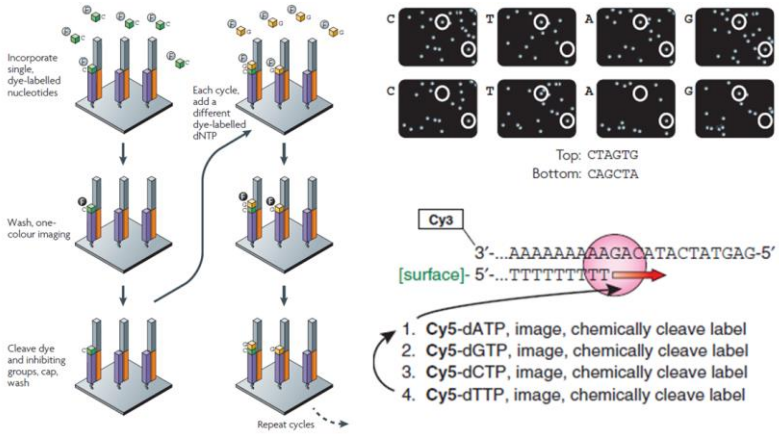


Figure 1.6. Schematic representation of the Helicos single molecule fluorescent sequencing (Shendure & Ji, 2008, Metzker, 2009).

1.1.2.5. Pacbio

PacBio RS platform was developed by Pacific Biosciences and its main characteristic is a single-molecule real-time (SMRT) sequencing system (Eid et al., 2009). For the first step of this genome sequencing, DNA is randomly fragmented and end-repaired. Then, 3’ adenine is added to this fragmented genomic DNA, facilitating the ligation of an adapter with a T overhang. Generally, single DNA oligonucleotide is used as the adapter because they forms an intramolecular hairpin structure. A single DNA polymerase is immobilized on the bottom of a reaction cell, which is called as zero-mode waveguides (ZMW). Then, nucleotides diffuse into the ZMW chamber. For the purpose of detecting incorporation and identify the base, each of the four nucleotides such as A, T, G and C is labeled with a different fluorescent dye having a distinct emission wavelength. The excitation illumination is directed to the bottom of the ZMW chamber, therefore, nucleotides held by the polymerase prior to incorporation emit an extended signal which identifies the base being incorporated. When the sequencing reaction begins, the tethered polymerase incorporates nucleotides with individually phospholinked fluorophores, to the growing DNA chain. Each fluorophore correspond to a specific base,(Eid et al., 2009). During the initiation of a base incorporation reaction, the fluorescent nucleotide is brought into the active site of polymerase and into proximity of the ZMW glass surface. At the bottom of the ZMV chamber, high-resolution camera continuously records the fluorescence signal of the nucleotide being incorporated. During the incorporation event, a

phosphate-coupled fluorophore is released from the nucleotide and this dissociation diminishes the fluorescent signal. Whilst the polymerase synthesizes a copy of the template strand, incorporation process of successive nucleotides could be recorded in a movie-like format.

.

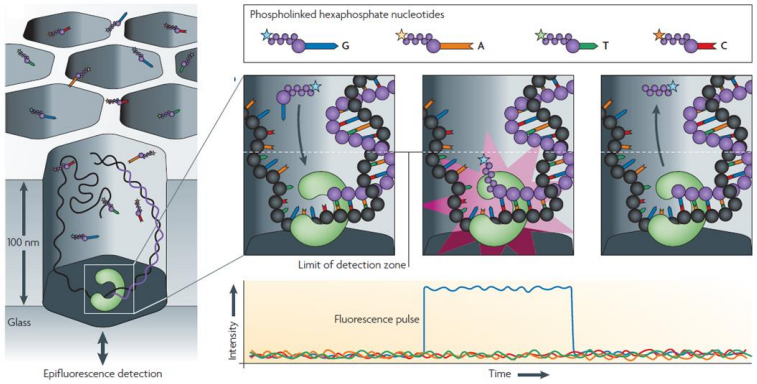


Figure 1.7. Schematic representation of PacBio RS system usinsingle-molecule real-time (SMRT) sequencing techniques (Metzker, 2009)

1.1.2.6. Nanopore Sequencing

Principle of nanopore sequencing is based on a mechanism that recording the current modulation of nucleic acids, which passes through a pore, could be used to distinguish the sequence of individual bases in the DNA chain. This Nanopore sequencing is expected to provide an effective solution to the problems of short read sequencing technologies and also make it possible to read the sequence of large DNA molecules in minutes without modification or preparation of samples. Expected characteristics of the nanopore sequencers are single molecule, amplification free, base detection without labels, long reads, low GC bias and scalable in data output. The basic principle behind the technology is tunneling of molecules through a pore that separates two compartments. Physical presence of the molecule passing through the pore causes a characteristic temporary change in the potential between the two compartments which allows for identification of the specific molecule. Two version of nanopore DNA sequencing are being developed, i.e., using the natural pore forming protein alphahemolysin (Stoddart et al., 2009) or manufactured solid state pores (Dela Torre et al., 2012). Oxford nanopore technologies (ONT) is one of the companies working on building nano-pore sequencing devices. They announced the early access release of their MinION system. This palm sized sequencing equipment can conduct real-time analysis of single molecules of DNA and RNA. However, specifics on read length, accuracy and run times are difficult to obtain.

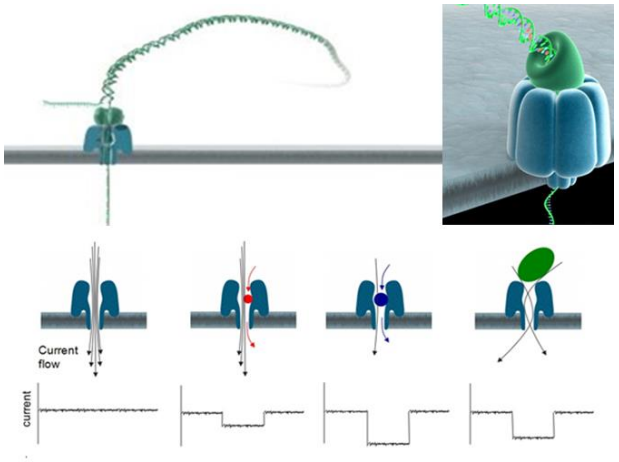


Figure 1.8. Schematic representation of Oxford Nanopore sequencing techniques. (Source: http://bit.ly//lltlndoreseq)

1.1.3 The application of next-generation sequencing

The next-generation sequencing has been applied for variety of research areas over the past several years. For example, Wheeler et al., reported comprehensive polymorphism and mutation discovery in individual human genomes through complete genome re-sequencing (2008). Van Tassel et al. conducted reduced representation approach for the discovery of large-scale polymorphism (2008). Next generation sequencing also could be used to find the targeted polymorphism and mutation. (Dahl et al., 2007, Frederickson et al., 2007, Porreca et al., 2007).Paired end DNA sequencing could provide a clue for Di Metagenomic sequencing revealed infectious and communal flora(Cox-Foster et al., 2007). Transcriptome sequencing enabled to find transcribed SNPs or somatic mutations, transcript annotation and quantitative analysis of gene expression and alternative splicing (Nagalakshmi et al., 2008, Sugarbaker et al., 2008, Wilhelm et al., 2008), discovery of inherited and acquired structural variation (Campbell et al., 2008, Chen et al., 2008). When applied to small RNA sequencing, these methods revealed the profiling of microRNA (Ordway et al., 2007, Korshunova et al., 2008, Lister et al., 2008, Morin et al., 2008). Ordway et al., determined patterns of cytosine methylation in genomic DNA using sequencing techniques of bisulfite-treated DNA (2007). Johnson and coworkers revealed Genome-wide mapping of protein-DNA interactions by Chromatin immune-precipitation sequencing (ChIP-Seq) (2007). Next-generation sequencing techniques have also been applied for nucleosome positioning(Johnson et al., 2007, Robertson et al., 2007, Schones et al., 2008)

and Molecular barcoding (Kim et al., 2007, Meyer et al., 2008). From now on, later application of the next-generation sequencing at research area is expected to grow fast. Through these various sequencing techniques, the unknown research fields will be revealed in only few years.

.

1.2 Metagenomics

Metagenomics can be defined as the study of the collective genomes in microbial community involving both cloning and analyzing the genomes without culturing the organisms in the community (Figure 1.9). Many microbial species are difficult to study in isolation. It is because they often fail to grow in laboratory culture, depend on other organisms for critical processes, or have become extinct, implying that these organisms represent only a fraction of the whole living or once-living organisms of research interest. Therefore, metagenomics has recently become a powerful tool for collecting information on microbial communities, by passing cultivation of individual species. Metagenomics also provides a relatively unbiased viewpoint not only of the community structure (species richness and distribution) but also of the functional and metabolic potential of a microbial community. (Hugenholtz & Tyson, 2008)

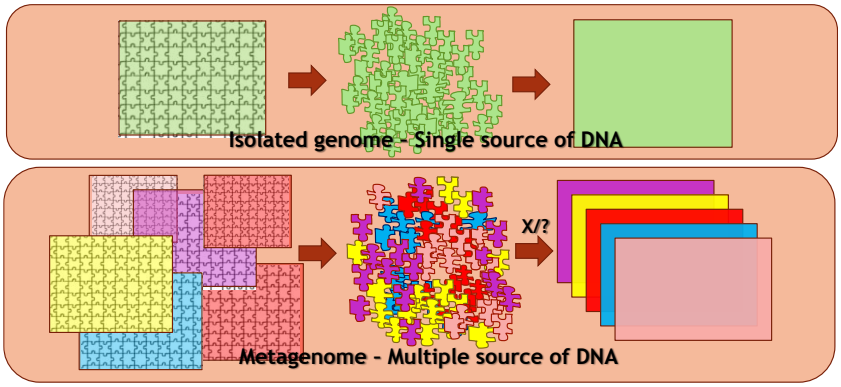


Figure 1.9. Concept of metagenomics from multiple source of DNA.

1.2.1. DNA Sequencing in Metagenomics

1.2.1.1. Needs for high-throughput sequencing

High-throughput sequencing is an indispensable tool in metagenomic study. The first metagenomic studies used massively parallel 454 pyrosequencing (Poinar et al., 2006). Three platforms of Ion Torrent Personal Genome Machine, the Illumina MiSeq or HiSeq and the Applied Biosystems SOLiD system have been applied to environmental samples (Rodrigue et al., 2010). These DNA sequencing techniques for generate shorter fragments compared to 1 generation of Sanger sequencing. In detail, both Ion Torrent PGM System and 454 pyrosequencing produce sequences reads about 400 bp. The read length of Illumina MiSeq system has been known to be 400~700 bp reads depending on conditions of paired end options. SOLiD produce read length of 25-75 bp (Schuster, 2007). These DNA sequencing techniques for generate shorter fragments compared to 1 generation of Sanger sequencing. In detail, both Ion Torrent PGM System and 454 pyrosequencing produce sequences reads about 400bp. The read length of Illumina MiSeq system has been known to be 400~700 bp reads depending on conditions of paired end options. SOLiD produce read length of 25-75bp. These read lengths range of high-throughput sequencing were significantly shorter than that of the typical Sanger sequencing read length (~750 bp). However, this limitation of relatively short read length can be compensated by the much larger sequence reads number. In year of 2009, it has been reported that pyrosequencing and Illumina

platforms generated metagenomes of about 200–500 megabases, and 20~50 gigabases, respectively. However, these outputs of metagenome have increased by orders of magnitude in recent years. An additional advantage of these high-throughput sequencing methods in metagenomics mentioned above is that they does not require the DNA cloning before sequencing, which remove one of the main biases and bottlenecks in environmental sampling for metagenomic study.

1.2.1.2. Shotgun sequencing approach in metagenomics

Progress in bioinformatics, improvements of DNA amplification, and the proliferation of computability have provide a new provision in the analysis of DNA sequences sampled from various environments As a result, method of shotgun sequencing could be easily adapted to metagenomic samples. This approach, which have been used to sequencing of many cultured microorganisms and the human genome, randomly cuts DNA sequences into many short sequences and then, reconstructs them into a whole consensus sequence, This means that though this shotgun sequencing , it was possible to reveal overall genes present in environmental samples. Clone libraries have been generally applied to facilitate this shotgun sequencing. However, advances in high throughput sequencing technologies make the situation that this cloning step is no longer necessary. Furthermore, greater sequencing data yields can be achieved without labor-intensive bottle neck step of clone library construction. Shotgun metagenomics can provide two kinds of information about of identification of organisms and possible metabolic process in the environmental community(Segata et al., 2013). The collection of DNA from an environmental condition is basically uncontrolled, which implies that the most abundant organisms in an environmental sample are also the most highly represented in the sequence data results. To achieve the high coverage which is necessary to fully resolving the genomes of under-represented community members, massively large samples should be required. On the other hand, the random nature of shotgun sequencing ensures that many of these organisms, which

would not be detected using traditional culturing approaches, will be represented by some small sequence segments (Tyson et al., 2004).

1.2.2. Estimation of Microbial Diversity by 16rRNA Gene Analysis

It was generally known that in a metagenomic study, one in every few thousand genes responds to a 16S rRNA gene. With 454 pyrosequencing, this typically translates to around 1.000 reads per picotiter plate which harbors partial 16S rRNA genes with sufficient quality and lengths for analysis of phylogeny. Depending on the length and region of the l 16S rRNA gene sequence, phylogenetic analysis can produce varying taxonomic depths. However, since the introduction of 454 pyrosequencing system, a substantial fraction of the respective reads enables a genus level assignment, and this situation is expected to improve in the future with increase of pyrosequencing read length. A limitation of pyrosequencing is that the number of obtained 16S rRNA genes with high quality might insufficient for a representative biodiversity estimation and unparticular not for lowly abundant taxa. Illumina platform does not have this limitation of insufficient number due to its much higher throughput. On the other hand, Illumina is troubled by its comparatively short reads which means that compromise between the depth and quality of the taxonomic assignments. Holmes and coworkers have successfully proposed dedicated analysis frameworks for clustering such data into operational taxonomic units (OTUs).(Holmes et al., 2012) Subsequently, representative sequences for OUT scan be mapped against a 16S rRNA reference tree for classification (Pruesse et al., 2007, Yilmaz et al., 2012). The advantage of this approach over 16S rRNA gene clone libraries is that no primers are involved and therefore, primer bias does not exist. Decides not obtaining high-quality

and full-length 16S rRNA gene sequences, the disadvantage, is that different taxa harbor different numbers of rRNA operons, distorting metagenomic 16S rRNA gene abundances. For instance, some Planctomycetes feature large genomes but only a single disjoint rRNA operon (Kotera et al., 2012), This inevitably lead to an underestimation of their abundance in relation to average-sized genomes with more rRNA operons. Analysis of metagenomic partial 16S rRNA genes provides a direct way to estimate a habitat's biodiversity which in the case of 454 pyrosequencing system usually can provide a resolution down to the genus level. The resulting information is essential for both identifying misclassifications in the taxonomic classification of other sequences and identifying taxa which could be passed over in the taxonomic classification process.

1.2.3. Functional Analysis in Metagenome

Metagenomic analysis entails functional annotation of the predicted genes through database search and comparison searches. This type of activity typically includes protein BLAST searches against databases of SWISSPROT, NCBI nr or KEGG (Kotera et al., 2012), HMMer searches against the Pfam (Punta et al., 2012) and TIGRfam (Selengut et al., 2007) databases, as well as tRNA and rRNA predictions (Lowe & Eddy, 1997) genes, signal peptides (Emanuelsson et al., 2007), transmembrane regions(Krogh et al., 2001, Liakopoulos et al., 2001), CRISPR repeats (Bland et al., 2007) and sub-cellular localization. Also, dedicated databases are available for unique functions, for instance the CAZY (Bhasin et al., 2005) database for carbohydrate-active enzymes, the TSdb (Zhao et al., 2011) database for transporters and the MetaBioMe (Zhao *et al.*, 2011) database for enzymes with biotechnological prospects. The resulting annotations could be used as a basis for functional data mining such as metabolic reconstruction. Dedicated systems such as WebMGA (Wu et al., 2011), IMG/M(Markowitz et al., 2006, Markowitz et al., 2008, Markowitz et al., 2012) and MG-RAST (Meyer et al., 2008, Glass et al., 2010, Mitra et al., 2011) have been developed for the purpose of supporting these metagenome annotation works. All of the three cases mentioned above have expanded annotation systems and continued to add useful features such as metagenome comparisons, biodiversity analysis, and taxonomic classification. Although automatic annotation, a substantial proportion of these annotations are erroneous or even incorrect. These limitation of bioinformatic methods was

expected to only be addressed by future high-throughput functional screening pipelines.

1.3 Metagenomic Approach in Domestic Animals

A commercial animal means any animals of a domestic species used for agriculture purposes and includes horses, cattle, hogs, goats, sheep, poultry and bee colonies but does not include wildlife as defined in The Wildlife Act, even if used for agricultural purposes. Metagenomics research in commercial animals can provide a new insight of environmental sustainability and productivity. That is to say, metagenomic approach enables us to better understand the relationship between relevant microbiomes and animal production efficiencies and well-being.

1.3.1 Bovine

Metagenomic approach in the bovine has been conducted in various topics of rumen microbiome, pathogenicity and fiber degrading enzymes. Elizabeth and coworkers profiled whole rumen metagenome using untargeted massive parallel sequencing technique method. Furthermore, they investigated whether rumen metagenome profiles were predictive of fecal metagenome profiles. DNA obtained from the samples was sequenced using Illumina GAIIx. When the reads were aligned to a reference rumen metagenome , their profiles were repeatable ($P<0.00001$) by cow regardless of location of rumen fluid sampling (Ross et al., 2012). Georgios and coworkers invested microbial diversity in bovine mastitic milk using pyrosequencing of metagenomic 16s rDNA. Discriminant analysis between mastitic and healthy dairy cows showed

that the groups of samples were most clearly different. Furthermore, the mastitis pathogens identified by culturing were generally among the most frequent organisms detected by pyrosequencing, and in some cases of *Escherichia coli*, *Klebsiella* spp. and *Streptococcus uberis mastitis* was identified to be the single most prevalent microorganism (Oikonomou et al., 2012). The complex rumen microbiome functions as an effective system which convert plant cell wall biomass to microbial protein, short chain fatty acids, and gases. Burulc et al. applied comparative metagenomics to investigate randomly sampled pyrosequence data from the fiber-adherent bovine rumen microbiome(Brulc et al., 2009). A comparison of the glycoside hydrolase and cellulosome functional genes showed that in the rumen microbiome, initial colonization of fiber appears to be by organisms possessing enzymes which targets the easily available side chains of complex plant polysaccharides and not the more recalcitrant main chains such ascellulose, especially. Rumen microbes of cow have specialty in in cellulosic plant material degradation, but most of these complex community are difficult to cultivate. . Metagenomic discovery of biomass-degrading genes and genomes from cow rumen revealed by Hess M, et al. They sequenced and analyzed metagenomic DNA recovered from plant fiber adherent microbes incubated in rumen to characterize biomass-degrading genes and genomes, As a result, they identified 27,755 putative carbohydrate-active genes and expressed 90 candidate proteins, whose 57% were determined to have enzymatic activity against cellulosic substrates (Hess et al., 2011). Wu et al. surveyed the rumen bacterial composition of pre-ruminant dairy calves

and cows and beef steers using the 16S rRNA gene pyrosequencing. They reported that the core rumen microbial community consisted of 8 phyla, 11 classes, 15 families, and 17 genera regardless of the rumen developmental status or breeds. Furthermore, principal component analysis and clustering showed that the bacterial communities in the rumen of pre-ruminant dairy calves, dairy cows, and beef steers were clearly distinguishable. That is to say, about 66% of phyla and 41% of OTUs in a typical rumen bacterial community showed difference in relative abundance between the developing and mature rumen. Greater abundance of and *Ruminococaceae, Fibrobacteraceae* in the beef steers rumen reflected the need for enhanced fiber-digesting capacity in beef cattle (Robert, 2012).

1.3.2 Chicken

It has been generally reported that chickens harbor unique and diverse bacterial communities including both human and animal pathogens. Besides, increasing public concern about antibiotics use in the poultry industry has affected the ways in which poultry producers are trying to improve birds' intestinal health. The complex microbiome of chickens the ceca plays an important role in nutrient utilization, growth and animal well-being. Qu et al. characterized the chicken cecal microbiome through comparison between pathogen-free chicken and one that had been challenged with *Campylobacter jejuni*. They showed that mobile DNA elements are a major functional component of cecal microbiomes, and contribute to both horizontal gene

transfer and functional microbiome evolution. Moreover, the metavirulomes of these microbiomes appeared to closely associate by host environment(Qu et al., 2008). In the perspective of poultry industry, it is desirable to minimize the performance variation which partially induced by gut microbiota. Singh et al. investigated metagenomic profile of fecal bacteria in birds with high and low feed conversion ratio (FCR) and identified microbial community linked to low and high FCR through employing high-throughput pyrosequencing of 16S rRNA genomic targets. As a result of this, they reported that the fecal microbial community of birds was predominated by three phylogenic taxa such as Proteobacteria (48.04 % in high FCR and 49.98 % in low FCR), Firmicutes (26.17 % in high FCR and 36.23 % in low FCR), Bacteroidetes (18.62 % in high FCR and 11.66 % in low FCR), as well as unclassified bacteria (15.77 % in high FCR and 14.29 % in low FCR),. These results suggested that a large portion of fecal microbiota was novel and could be linked to currently unknown functions(Singh et al., 2012). Tang et al. carried out both high-throughput sequencing of 16S rRNA gene amplicons and metaproteomics analysis of fecal samples to determine composition of microbial gut and protein expression. As a result of this, 16 rRNA gene sequencing analysis successfully identified *Clostridiales*, *Bacteroidaceae*, and *Lactobacillaceae* species as the most dominant bacterial species in the gut. Among the most frequently identified 3,763 proteins from metaproteomics, each number of proteins belonged to *Lactobacillus* spp., *Clostridium* spp. and *Streptococcus* spp. were 380, 155 and 66, respectively (Tang et al., 2014).

1.3.3 Goat

Goat is an economically important small ruminant having remarkable capacity to digest different type of feeds and fodder. In more detail, goat's rumen contains a rich resources of microbes for the degradation of various cellulosic biomass. Jakhesara et al. performed a comparative metagenomic analysis of rumen samples from 4 healthy goats randomly selected from population to obtain a detailed characterization of the goat rumen microbiome using a tannin free diet. They have reported selective increase in the no. hits in case of Bacteroidetes, *Clostridia*, *Proteobacteria* (*Alpha, Beta, Gamma, and Delta*), and *Actinobacteria* and decreased hits in case of *Metazoa*, *Euryarchaeota* and *Cyanobacteria* upon tannin treatment, which suggests harmful and beneficial effect of tannin on rumen microbes (Jakhesara et al., 2010). Zhou and coworkers extracted metagenome DNA from meat goat's rumen fluid and conducted a paired-end sequencing analysis followed by assembly and annotation. In the annotated database, full or partial length gene sequences for beta-exoglucanases, beta-glucosidases and cellulases were identified. Some of the interesting genes are cloned from the same metagenome DNA extract using PCR approach, and their potential use in the production of bioenergy industry are being investigated(Zhou, 2014). Wang et al. investigated the genetic diversity of xylanases which belong to two major glycosyl hydrolase families of GH 10 and 11 in goat rumen contents by analyzing the amplicons produced using two degenerate primer sets. Phylogenetic analysis showed that all of GH 10 xylanase sequences fell into seven clusters, and 88.5%

of them were related to xylanases from Bacteroidetes. In addition, five clusters of GH 11 xylanase sequences were identified. Among them, 85.7% were determined to be related to xylanases from Firmicutes, and 14.3% were determined to be related to those of rumen fungi, respectively (Wang et al., 2011). Souto and coworkers constructed a small insert metagenomic library using environmental DNA obtained from the solid portion of Moxotó goats' rumen, which is known to be a breed of goats native to the semi-arid region of Brazil. As a result of this, two cellobiohydrolase and three β-glucosidase clones that presents strong phenotypes had their enzymatic activities confirmed by retransforming their plasmidial DNA into *Escherichia coli*. In addition, the clones with cellobiohydrolase activity also demonstrated similarity to cellulases from the microbiota of other ruminant (SOUTO et al., 2011).

This chapter will be published in elsewhere
as a partial fulfillment of Sooyeon Lim's Ph.D. program.

Chapter 2. Microbial community structures of rumen solid and liquid in cattle.

2.1 Abstract

Sequencing the genomes of rumen microbes, determining the role of the genes and identifying the potential applications are the great deal for researchers to understand the microbiology of the rumen. Another complication is quantification of rumen microbes in the laboratory due to the diversity and density of microbes in the rumen. In this study, the microbial community structure of rumen solid and rumen liquid of cattle rumen was analyzed using high-throughput pyrosequencing of 16S rRNA gene PCR amplicons and a subsequent bioinformatics pipeline. A 16S rRNA gene clone library identified abundant communities belonging to specific bacterial groups in the rumen. The diversity results suggested that the specific bacterial groups was found in both samples with a slight difference. Bacteroidetes, Firmicutes and Fibrobacteria were present in higher concentrations in rumen solid than in rumen liquid, indicating their major role in the degradation of plant fiber. Other groups identified include Proteobacteria, which are responsible for playing a greater role in rumen metabolism; *Mollicutes*, which metabolizes imported sugars to short chain fatty acids; and *Prevotellacea*, which are crucial for the breakdown of proteins. These biological function of identified microbial groups mentioned above have also coincided with other previous studies.

2.2 Introduction

Mammalian guts harbor a diverse microbial ecosystem which benefits their host in many aspects including digestion of unutilized energy substrates (Leschine, 1995), stimulation of cell growth, repression of pathogen growth (Nuijens et al., 1996), protection against epithelial cell injury (Nuijens et al., 1996, Kudva et al., 2014) , regulation of host fat storage (Backhed et al., 2004) stimulation of intestinal angiogenesis(Stappenbeck et al., 2002) , enhancement of host immune functions(Ross et al., 2010) and protection against disease(Li, 2003). The rumen microbial system mainly consists of obligate anaerobic microorganisms such as bacteria, archaea, fungi and protozoa. Gut flora provide the important metabolic function of gleaning energy from indigestible dietary polysaccharides(Leschine, 1995) and the subsequent absorption of short-chain fatty acids(Hijova & Chmelarova, 2007). Gut bacteria also plays an essential role in synthesizing vitamins as well as metabolizing bile acids(Jones et al., 2008), sterols, and xenobiotics (Cummings & Macfarlane, 1997, Hijova & Chmelarova, 2007, Jones et al., 2008). An imbalance in the microbial community in mammalian gut may lead to metabolic syndromes, inflammation, weight loss or gain, and obesity(Ott et al., 2004, Manichanh et al., 2006, Frank et al., 2007) While the microbial flora plays a fundamentally major role in health and disease (Eckburg et al., 2005, Schogor et al., 2014), this ecosystem remains incompletely characterized and its microbial diversity poorly defined (Hooper & Gordon, 2001). It is essential to evaluate the microbial diversity in

the bovine rumen for understanding its important roles in the animal's overall health, productivity, and well-being. Sequence-based metagenomic analysis has revealed the core metabolic functions of the mammals gut microbiota(Tajima et al., 1999) and determined groups of communities, including many previously uncultured ones, in ruminant samples (Koike et al., 2003). Metagenomic investigation is comprised of isolation of total DNA from entire microbial community, cloning and library construction, sequencing the clones, assembling the sequence into contigs and scaffolds, analysis of complete microbial communities, and interpretation of results in order to understand the microbe-environment interactions. Metagenomic approaches have been widely used to identify entirely new classes of genes for new or known functions. For instance, isolation of novel biocatalysts from environmental samples (Daniel, 2005, Lorenz & Eck, 2005), identification of cellulose genes in metagenomic libraries (Rees et al., 2003) from various environmental samples, novel hydrolase (Ferrer et al., 2005) and a multifunctional hybrid glycosyl hydrolase (Palackal et al., 2007) from bovine rumen microflora and protein domain collocation within cellulose genes from microbes of goat rumen (Lim et al., 2013) have been performed. Genome sequences of uncultured bacteria in rumen have begun to be used along with functional genomics tools such as transcriptomics, proteomics and metabolomics to attain a greater understanding of environmental microbiology. The complex microbial community harbored in rumen have properties which play key roles in providing various nutrients to the host. A key challenge in studying the rumen-microbial ecosystem is

identifying the symbiont microbes in the rumen; 99% of microbes cannot be isolated or cultured due to the limitations of culture-based methods in microbiology (Hugenholtz, 2002). Standard culture-based techniques are unable to capture the true microbial diversity(Degnan & Ochman, 2012) in any ecosystem. However, metagenomic approaches provide a global microbial gene pool without the need to culture microbes (Sjöling et al., 2007). In this study, it has applied metagenomic approaches for isolation of the complete genome from the entire microbial communities in cattle rumen. The microbial profile in the rumen environment was characterized through massive pyrosequencing and bioinformatics.

2.3 Materials and methods

Sample collection and DNA extraction

Rumen fluid was collected from a 1-year-old Korean native cattle and Saanen hybrid raised on Timothy (Phleumpratense) hay at a private cattle farm in the Cheonan, Korea area and slaughtered at a local slaughter house. Rumen fluid was filtered through four layers of cheesecloth. Genomic DNA was isolated from rumen fluid using the Wizard Genomic DNA Purification Kit (Promega, U.S.) according to manufacturer protocol. Gel electrophoresis was performed with 1% agarose gel at 50V for 2 hours to check both quality and quantity of isolated genomic DNA. About 1L of rumen digesta was collected in the morning after 90 minutes of feeding, and was squeezed and filtered through a coarse muslin cloth to separate rumen solid and filtrate. The filtrate was labeled as rumen liquid after filtration using 20 μ nylon mesh for removal of fungi and debris. The filtrate was aliquoted in centrifugation tubes and stored at -80℃ for further study. The undigested fiber was named rumen solid, aliquoted in sterile tubes, and immediately frozen in liquid nitrogen. Both rumen liquid and rumen solid samples were used for analysis of the microbial community in cattle rumen through pyrosequencing (Figure 2.1). Metagenomic DNA was extracted using the hexadecylmethyl ammonium bromide and sodium dodecyl sulfate (CTAB-SDS) method (Zhou et al., 1996). 5g samples were homogenized by vortexing and then incubated at 37℃ for 2 hours with continuous shaking at 225 rpm in 13.5 mL of extraction buffer (100 mM Tris-

HCl pH 8.0, 100 mM sodium EDTA pH 8.0, 100 mM sodium phosphate pH 8.0, 1.5 M NaCl, 1% CTAB) and 100 µL of proteinase K (20 mg/mL). After addition of 6 mL of 10% SDS, samples were incubated at 37℃ for 2 hours. Supernatants were collected by centrifugation at 2200 × *g* for 30 min at 15℃ and then filtered with 4-folded gauze. Lysates were extracted twice with chloroform and isoamyl alcohol (24:1). Metagenomic DNA was precipitated with 0.6 volumes of isopropanol, re-suspended in TE buffer (pH 8.0) to give a final volume of 500 µL, and stored at -20℃ before use.

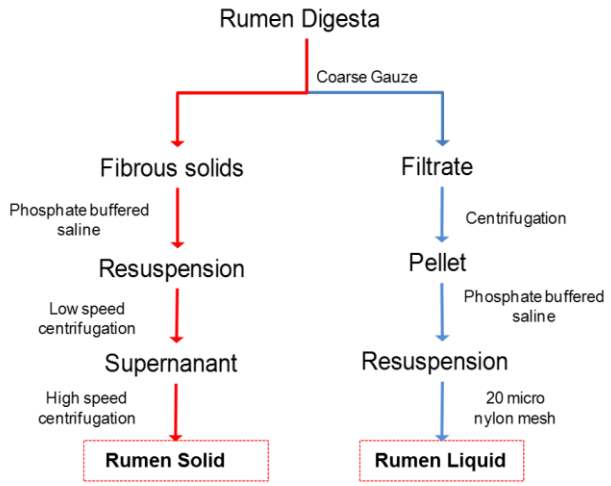


Figure 2.1. Schematic diagram of sampling from bovine rumen.

Bacterial 16S rRNA amplification

Full length 16S rRNA was amplified from universal barcoded primers of 9F (5'-CCTATCCCCTGTGTGCCTTGGCAGTC–TCAG–AC–AGAGTTT GATCMTGGCTCAG-3'; underlining sequence indicates the target region primer) and reverse primers 541R (5'-CCATCTCATCCCTGCGTGTCTCCG AC-TCAG-X-AC-ATTACCGCGGCT GCTGG-3'; 'X' indicates the unique barcode for each subject) were used. PCR reactions were carried out in a thermo cycler (MJ Research, Reno, USA) under the following conditions: initial denaturation at 94℃ for 5 min, followed by 25 cycles of denaturation at 94℃ for 30 sec, annealing at 60℃ for 30 sec, and elongation at 72℃ for 1 min 20 sec. The amplified 16S rRNA was purified using resin spin column and the one μg of each amplified products were mixed and subjected to pyrosequencing. Pyrosequencing was carried out by Macrogen Inc. (Seoul, Korea) using a 454 Genome Sequencer FLX Titanium (Roche, Basel, Swiss) according to manufacturer's instruction.

Barcode processing

Both nucleotide sequences and converted PHRED (Ewing and Green, 1998) quality scores were generated using 454 Genome Sequencer (model FLX Titanium). Sequencing reads from the different samples were separated by 4 to 8 nucleotide long barcodes which were included at both forward and reverse

primers, with barcodes and 2 bp linker (AC) trimmed. Separated reads with 50 or more bp were saved in fastaq format.

Primer trimming

Both forward and reverse primers were trimmed from each sequencing read to avoid primer inclusion. Using the Myers & Miller pairwise sequence alignment algorithm (Myers & Miller, 1988) to detect the exact position of PCR primers at both 5' and 3' ends of each sequencing read. Sequences read from the reverse primer were transformed into their inverted complementary, so sequences in the resultant fastaq file were all in the same direction (5'→3').

Removal of short and low quality sequencing

Sequencing reads containing ambiguous base calls (Ns) were reported to contain significant amount of sequencing errors in 454 GS-20 machine. However, according to our evaluation based on 454 FLX Titanium, sequences with one N of 454 showed high sequence quality comparable to sequences with no Ns. Therefore, excluding sequencing reads with two or more Ns and short reads (less than 200 base calls with Q<20) were removed from the dataset.

Removal of non-specific PCR amplicons

Although PCR is designed to amplify specific products, non-targeted regions of template DNA may amplify at a low frequency. To remove these

reads, all reads are searched against a 16S rRNA gene database containing representatives of Bacteria and Archaea using the BLASTN program (Altschul et al., 1997). If a sequencing read shows no match (BLAST expectation value of > e^{-5}), it is considered non-specific amplicon and removed from subsequent analyses.

Assembly/clustering of sequencing reads

Individual sequencing reads were clustered into groups based on sequence similarity. Developing an algorithm to assemble sequence reads while ignoring errors at homopolymeric positions. To achieve this, all sequences were initially "condensed"; any homopolymeric sites were converted to single base calls (e.g. AAATTCGG to ATCG). Next, each pair of condensed reads was aligned using the Myers & Miller global alignment algorithm (Myers & Miller, 1988) and merged into a group if they showed no mismatches. In this process, indels represented as gap in the alignment and not considered. Once grouping was achieved, consensus sequences for each group were calculated from the original sequences with the majority rule. Resultant groups were further merged into fewer clusters if they showed four or less mismatches. At each step, consensus (=contig) sequences were constructed for each group of reads under the basis of majority rules. Using this iterative procedure, pyrosequencing reads were clustered into contigs in which sequence variation among its members is less than 1%.

Taxonomic assignment of individual sequencing reads and contigs

Developing a database, named EzTaxon-extended, which contains 16S rRNA data of representative bacterial phenotypes. It constists of all entries of the EzTaxon database (http://www.eztaxon.org (Chun et al., 2007), with 16S rRNA gene sequences of type strains and representative phylotypes of either cultured or uncultured entries in the GenBank public databases. Representative phylotypes are designed as artificial taxonomy', and for each of these an artificial specific epithet was given. For example, epithet AY510255 was assigned to the GenBank sequence entry AY510255, which represents an artificial taxonomy. Similarly, artificial taxonomic ranks higher than genus are assigned where appropriate. In addition to the assignment of representative sequences to type strains of artificial taxonomy, creating a new hierarchical taxonomic system for known bacterial 16S rRNA gene sequences. In this new system, it assigned over 20,000 sequences with complete taxonomic hierarchy, i.e. genus, family, order, class, phylum, domain. Individual pyrosequencing reads or contig sequences generated by the assembly pipeline were compared to sequences in the EzTaxon-extended database using BLASTN. The five sequences with the best hit scores were then scanned for pairwise sequence similarity using global sequence alignment, as described by Chun et al(Chun et al., 2007). The best match, defined by the highest sequence similarity, was used for taxonomic assignment of query sequences. The criteria for taxonomically categorizing Individual sequence was as follows, where indicates similarity: genus ($x > 94\%$), family ($94 > x > 90\%$), order ($90 > x > 85\%$), class ($85 > x >$

80%) and phylum (80 > x > 75%). If similarity did not meet the cutoff, the read was assigned as unclassified group.

Calculation of taxonomy composition richness and diversity indices

Multiple alignment of individual pyrosequencing reads was generated using the ClustalW-mpi program (Li, 2003). Diversity, taxonomy richness indices (Table 2.1), and the rarefaction curve were calculated using the DOTUR program(Schloss & Handelsman, 2005). The cutoff value for assigning a sequence to the same group was equal to or greater than 97% similarity.

Table 2.1 Number of validated sequences and comparison of phylotype coverage and diversity estimation at 97% similarity in operational taxonomic unit (OTUs), estimated OTU richness (ACE and Chao 1), diversity indices (Shannon and Simpson) and Goods Lib. Coverage in rumens Solid and Liquid.

Sample	Valid reads	OTUs	Ace	Chao1	Shannon	Simpson	Goods Lib. Coverage
Rumen Solid	7538	2504	5679.56	4207.32	7.17	0.0016	0.83
Rumen Liquid	5750	2767	5965.87	5339.62	7.46	0.0011	0.70

2.4 Results

Pyrosequencing data sets

The microbial community profile in the rumen environment were identified using metagenomic analysis with both pyrosequencing and bioinformatics. A total of 13,288 sequence reads were generated from the two samples after processing. The average length of processed sequence reads was 448 bases (figure 2.2). In addition, rarefaction curves (Figure 2.3) of all samples had reached the curvilinear phase, implying that the sampling was sufficient to obtain total diversity. The isolated genes from rumen liquid were obtained 7,538 number of taxonomic composition in rumen liquid sample of rumen and 5,750 taxonomic composition in rumen solid.

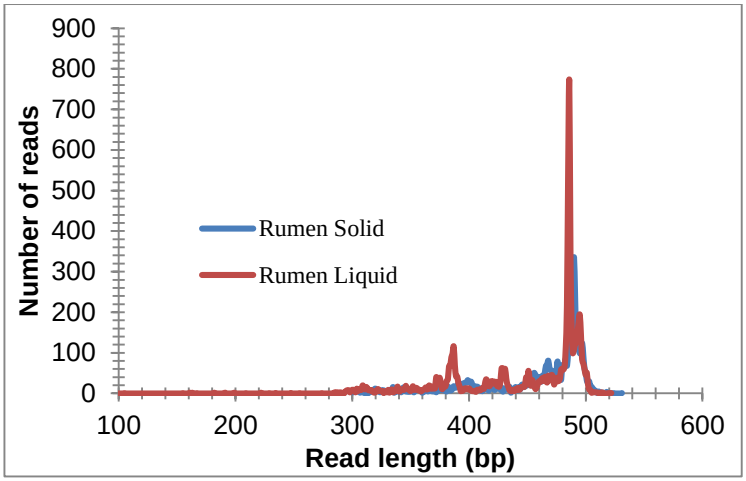


Figure 2.2. Average length of sequence reads

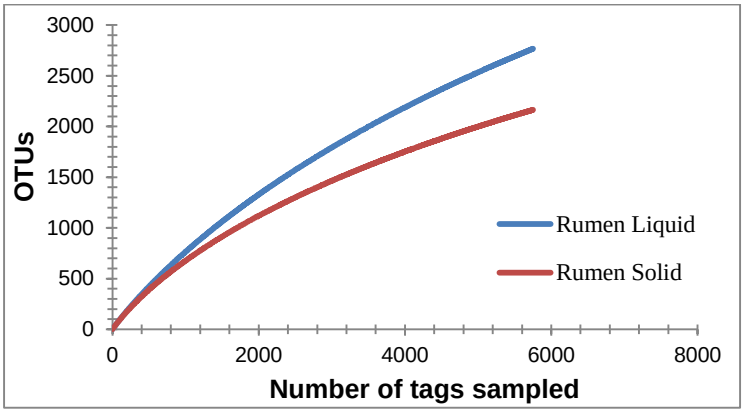


Figure 2.3. Rarefaction analysis of pyrosequences.

Community profile analysis at phylum level

The bacterial structure of rumen liquid and rumen solid was depicted in phylum level (Figure 2.4a). The rumen liquid sample was consisted of 17 phyla and its major groups were identified to be Bacteroidetes (75%), Proteobacteria (11%), Firmicutes (8%), Cyanobacteria (2%) and Spirochaetes (1%). On the other hand, the total 19 phyla were identified (Figure 2.4a) in the sample of rumen solid, whose main groups were Bacteroidetes (59%), Firmicutes (19%) and Proteobacteria (6%), Fibrobacteres (4%), Spirochaetes (3%), Lentisphaerae (3%), Tenericutes (2%) and Cyanobacteria (2%) (Figure 2.5a). Minor phylum groups with less than 1% were summarized in Appendix 1. Planctomycetes (0.23%) and Chloroflexi (0.02%) were only identified in rumen solid. Especially, Fibrobacteres (4%) whose portion increased more than 10 times compared to the rumen liquid Figure 2.5a. This results indicates that the fiber degrading bacterial community (Ransom-Jones et al., 2012) were present in higher titer in rumen solid than rumen liquid .

In this study, the phylum includes Firmicutes, Proteobacteria, Spirochetes and Bacteroidetes and unknown bacteria were predominant in the ruminal epithelial tissue-associated bacterial community(Tajima et al., 1999). The similar phyla were detected from the rumen of sheep and heifers and reported that the dominant bacteria in epimural community in ruminants(Sadet-Bourgeteau et al., 2010). The phylum of Bacteroidetes and Firmicutes were present in rumen samples for playing a major role in the degradation of

polysaccharides (Forsberg et al., 1997) and Proteobacteria were responsible for playing a greater role in rumen metabolism (Kang et al., 2013).

Community profile analysis at class level

In class level, a total of 34 taxons were identified in the sample of rumen liquid and 35 taxons in the sample of rumen solid (Figure 2.4b). Both the samples showed some of the similar community of Bacteroidetes (75%) in rumen liquid and Bacteroidetes (59%) in rumen solid, *Clostridia* (10%) in rumen liquid and 18% in rumen solid, *Gamma Proteobacteria* (7%) in rumen liquid and 5% in rumen solid, *Spirochaetes*_c (1%) in rumen liquid and 3% in rumen solid and represented the dominant class (Figure 2.4b). Some other taxons such as *Lentisphaera_c* (3%), *Mollicutes* (2%), *AB355089_c* (2%), *Erysipelotrichi* (2%) and *AlphaProteobacteria* (1%) were found in major community in rumen solid (Figure 2.5b), which implies that these class proliferated in the environment of hays in the bovine rumen. Minor classes with less than 1% community share were summarized in Appendix 2. *Bacteroides*, *Spirochaetes* and *Lentisphaera* were present in higher proportion in rumen for playing a function of cellulose degradation(Leschine, 1995). The class of *Mollicutes* were largely identified, which metabolize the imported sugars to short chain fatty acids in rumen (Turnbaugh et al., 2008).

Community profile analysis at order level

In the rumen liquid, a total of 62 orders were identified and its major groups were *Bacteroidales* (74%), *Aeromonadales* (9%), *Clostridiales* (6%), and *AB355089_o* (2%), respectively (Figure 2.4c). In the sample of rumen solid founded 69 different orders namely, *Bacteroidales* (57%), *Clostridiales* (16%), *Aeromonadales* (5%), *Fibrobacterales* (4%), *Spirochaetales* (3%), *AB185535_o* (2%), *Bacteroidia_uc* (2%), etc., represented the dominant order groups (Figure 2.4c). Other minor order groups in each sample were summarized in Appendix 3. Some of the orders such as *Aminobacterium_o, Planctomycetales, Lactobacillales, Anaeroplasmetales, Rhizobiales, Elusimicrobiales, Verrucomicrobiales, Flavobacteriales, Anaerolinaeles, Opitutae_uc, Verrucomicrobia_uc_o, Planctomycetacia_uc*, and *Actinobacteria_c_uc* were only detected in rumen solid and were responsible for plant material degradation into sugars by hydrolytic enzymes. Among them, some specific community of *Fibrobacterales*, Spirochaetales, *AB185535*_o and *Bacteroidia_uc* were largely present compared to the community of rumen liquid (Figure 2.5c) for the degradation of cellulose into short chain sugars, amino acids and fatty acids.

Community profile analysis at family level

In family level, a total of 110 groups were identified in the sample of rumen liquid and 120 in rumen solid The dominant family such as

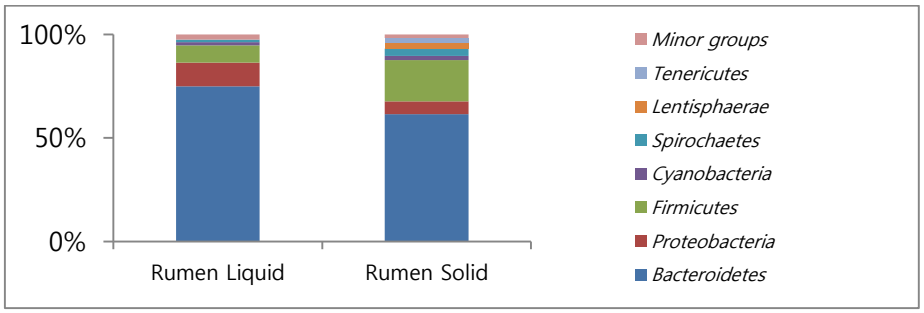
Prevotellaceae (67%), *Succinivibrionaceae* (7%), *Bacteroidales_uc* (3%), *AB185751_f* (2%), *Lachnospiraceae* (2%), *Clostridiales_uc* (2%), *Veillonellaceae* (2%) were determined in the community of rumen liquid, (Figure 2.4d). On the other hand, the community of *Prevotellaceae* (37%) was the most dominant in rumen solid but it is lower compared to rumen liquid. Some of the family in rumen solid includes, *Lachnospiraceae* (8%), *Bacteroidales_uc* (6%), *ParaPrevotella_f* (5%), *Clostridiales_uc* (4%), *Fibrobacteraceae* (3%), *Succinivibrionaceae* (3%), *Spirochaetaceae* (2%), *Ruminococcaceae* (1%) and *Erysipelotrichaceae* (1%) were largely increased compared to those portions in the rumen liquid. Minor groups were took the account of 12% in rumen liquid and 15% in rumen solid and listed in Appendix 4. This result indicate that *Prevotellaceae* was the most prevalent family identified in rumen and helped for the breakdown of protein (Wallace & McKain, 1991, Walker et al., 2003) and carbohydrates.

Community profile analysis at genus level

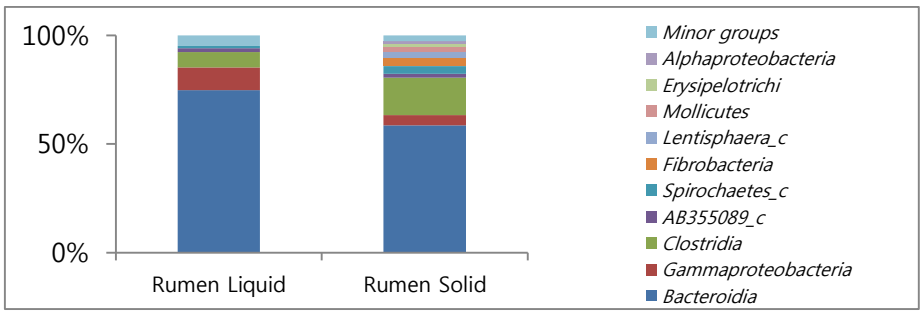
As clearly shown in Figure 2.4e, a total of 193 genus were identified in the rumen liquid community. Among them, only a few genera such as *Prevotella*, *Succinivibrio*, *Prevotellaceae_uc*, *Bacteroidales_uc_g*, *AB185751_g*, *Clostridiales_uc_g*, *Lachnospiraceae_uc* represent major microbial groups with each percentage of 62%, 5%, 5%, 3%, 2%, 2% and 1%, respectively. In rumen solid, *Prevotella* (30%) was determined to be the most

dominant genus among 216 genera, but its portion was largely lower than that of rumen liquid community. Additionally, various groups of *Prevotellaceae_uc* (7%), *Bacteroidales_uc_g* (6%), *Clostridiales_uc_g* (5%), *Lachnospiraceae_uc* (4%), *AF018497_g* (4%), *Fibrobacter* (3%), *Succinivibrio* (2%), *Bacteroidia_uc_g* (2%), *ParaPrevotella_f_uc* (2%), *Butyrivibrio* (1%), *AB009195_g* (1%), *AB185751_g* (1%) and *Clostridia _uc_g* (1%) were identified to be main genera (Figure 2.4e and 2.5e). The minor genera with less than 1% portion was summarized in Appendix 5. The population of the genus *Prevotella* was the predominant among the total bacteria and other genus were several fold smaller than *Prevotella*. The genus *Prevotella* responsible for protein degradation, *Fibrobacter* for fiber degradation, unclassified *Bacteroidales*, *Ruminococcaceae* and *Clostridiales* for biohydrogenation(Huws et al., 2011).

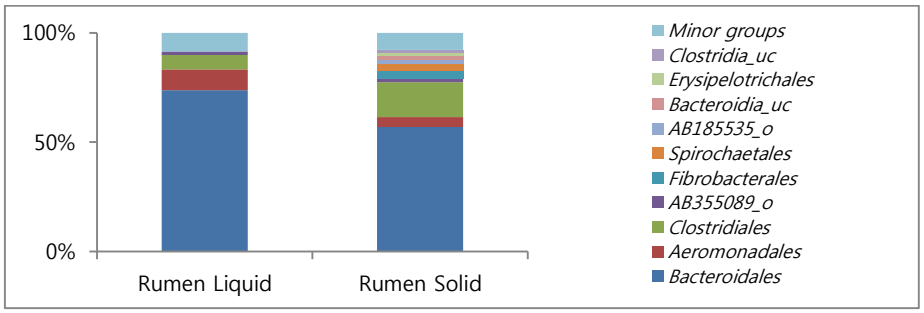
(a)


100%
50%
0%
Rumen Liquid
Rumen Solid
Minor groups
Tenericutes
Lentisphaerae
Spirochaetes
Cyanobacteria
Firmicutes
Proteobacteria
Bacteroidetes

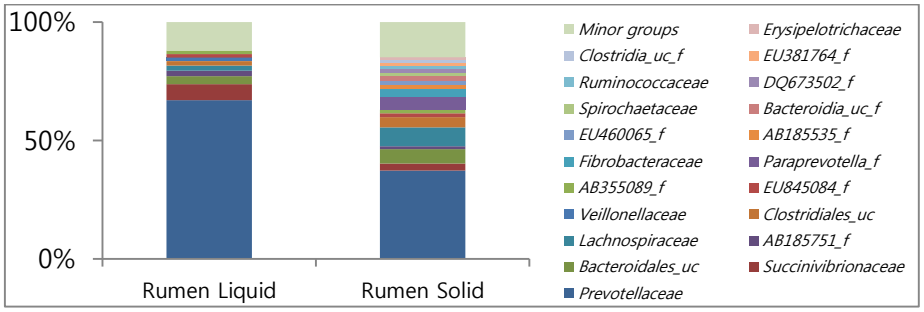
(b)


100%
50%
0%
Rumen Liquid
Rumen Solid
Minor groups
Alphaproteobacteria
Erysipelotrichi
Mollicutes
Lentisphaera_c
Fibrobacteria
Spirochaetes_c
AB355089_c
Clostridia
Gammaproteobacteria
Bacteroidia

(c)


100%
50%
0%
Rumen Liquid
Rumen Solid
Minor groups
Clostridia_uc
Erysipelotrichales
Bacteroidia_uc
AB185535_o
Spirochaetales
Fibrobacterales
AB355089_o
Clostridiales
Aeromonadales
Bacteroidales

(d)



(e)

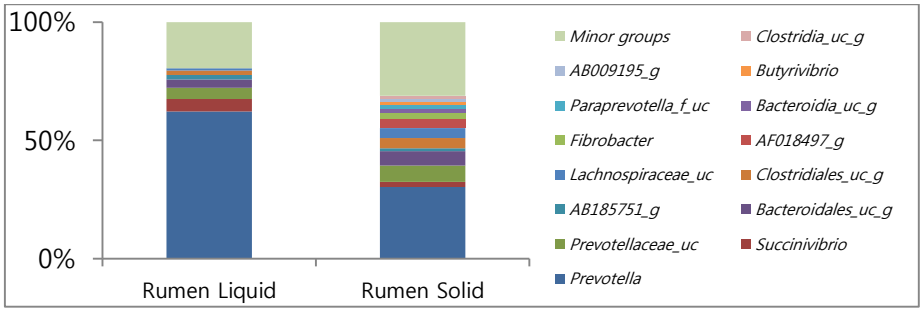


Figure 2.4. Phylogenetic distribution of each microbial community in (a) phylum, (b) class, (c) order, (d) family and (e) genus level, respectively

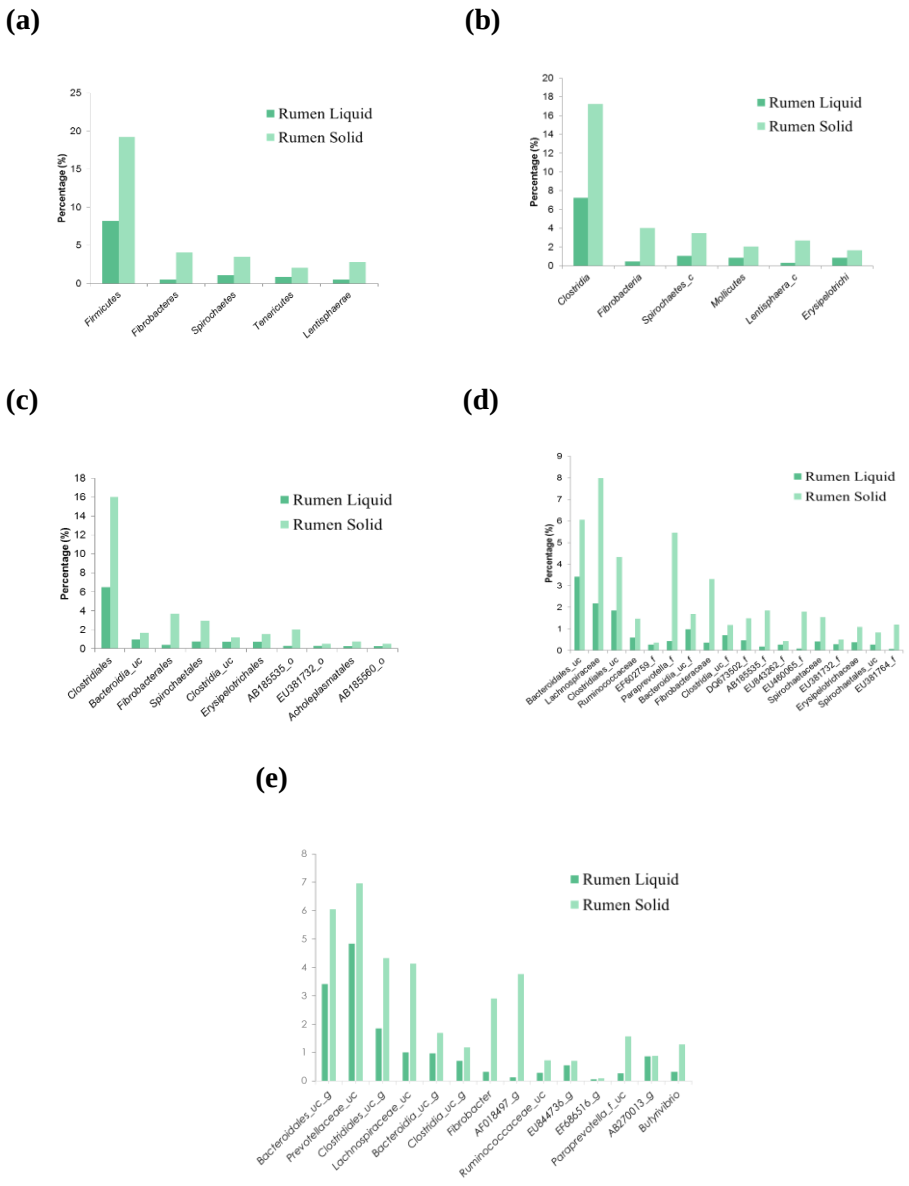


Figure 2.5. Comparison of distribution of major phylogenetic groups in each rumen niche (a) phylum, (b) class, (c) order, (d) family, (e) genus level, respectively

2.5 Discussion

The trait of microbial community profile in the rumen environment were identified using metagenomic analysis with both pyrosequencing and bioinformactics. A 16S rRNA gene clone library revealed that the community belonging to Bacteroidetes, Firmicutes, Proteobacteria, Spirochetes and Fibrobacter were the most abundant bacterial groups in the rumen. Especially in the phylum level, Firmicutes, Fibrobacteres, Spirochaetes, Tenericutes, and Lentisphaerae increased their ratio in the rumen solid more than two times compared to the rumen liquid. On the other hand, Fibrobacteres showed drastic increase in the ratio (8 times), which implies the presence of the fiber digesting bacterial groups with higher portion in rumen solid. In the Bacteroidetes phylum, the majority of sequences were assigned to class *Bacteroidia, Succinivibrio* and *Prevotella* was the most predominant genus. *Prevotella* is generally in high abundance in the rumen generally and represented by the species *Prevotella ruminicola* and *uncultured Prevotella*. Other phylum such as *Proteobacteria, Firmicutes, Cyanobacteria* and *Spirochaetes* were the dominant and played the major role in plant fiber degradation. Pyrosequencing analyzed the microbial community in the class level, *Clostridia*, *Gamma Proteobacteria* and uncultured groups were present in higher proportion for the degradation of protein and carbohydrates (Forsberg et al., 1997). Additionally, I found the major genera of *Prevotella, Fibrobacter, ParaPrevotella* and

Clostridia were playing the major role in cellulose degradation and ruminal biohydrogenation (Huws et al., 2011).

This chapter will be published in elsewhere
as a partial fulfillment of Sooyeon Lim's Ph.D. program.

Chapter 3. Analysis of bacteria community change in chicken feces during the rearing period using next generation sequencing

3.1 Abstract

Poultry contamination can be largely attributed to the presence of chicken feces during the production process. Fecal contamination is often found in raw chicken products sold for human consumption. Quantitatively analyze the fecal microbial community of chickens using next generation sequencing (NGS) techniques in this study. Fecal samples were collected from 30 broiler chickens at two time points- days 1 and 35 of development. 454 pyrosequencing was conducted on 16S rRNA extracted from each sample, and microbial population dynamics were investigated using various automated bioinformatics pipelines. Diversity of the microbial community at the genus level increased during the five-week growth period. Despite this growth, only a few dominant bacteria groups (over 80%) were identified in each fecal sample, with most groups being unique and only a few shared between samples. Population analysis at the genus level showed that microbial diversity increased with chicken growth and development. Classification and phylogenetic analysis of highly represented microbes (over 1%) clearly showed high levels of sequence similarity between groups such as Firmicutes and Proteobacteria. These results suggest that the chicken fecal excreted microbiome is a dynamic system with a differentiated population structure that harbors a highly restricted number of higher taxa.

3.2 Introduction

Commercial poultry products have been implicated as a leading cause of human infections(Waldroup, 1996, Vandeplas et al., 2010), and further research and accurate information on intestinal pathogens is needed for optimization of the poultry production process and contamination detection in consumable goods. Previous studies have investigated microbiota present in the intestinal tract of chickens; for example, species-specific fecal microbial sequences have been identified using a metagenomic approach(Lu et al., 2007). Another study investigated microbiota in the cecum of broiler chickens using temperature gradient gel electrophoresis. Analyses identified 50 phylogenetic groups or subgroups of bacteria, with approximately 89% of the sequences representing four phylogenetic groups- *Clostridium leptum*, *Sporomusa sp.*, *Clostridium coccoides*, and *Enterics*(Zhu et al., 2002). Recent development of commercially available next-generation sequencing (NGS) platforms has made more detailed profiling of the microbiota community structure possible. For instance, one study analyzed the distribution of genera in the microbiota in chicken caecum and identified 197 Operational Taxonomic Units (OTU) through next generation sequencing of 16S rDNA libraries(Nordentoft et al., 2011). Similarly, 454-pyrosequencing was used to investigate the distribution and composition of antibiotic-resistant bacteria in chicken manure and fertilized vegetables and revealed that chicken manure predominantly harbored Firmicutes, Bacteroidetes, Synergistetes, and Proteobacteria(Yang et al., 2013).

Although extensive studies have been conducted on this topic, most of which focused on the cecum, only few direct characterizations of the microbial community present in chicken feces have been performed. The growth periods of chickens are relatively short compared to other livestock animals such as pig and cow, which makes understanding the dynamics of its microbial flora crucial. Aim primarily to characterize the microbiome present in chicken feces during development using next generation sequencing platforms and bioinformatics tools in this research. And also it assess the microbial community profile quantitatively in terms of community structure at the taxonomic level, major microbial groups, and the phylogenetic correlation of the profile.

3.3 Materials and methods

Data analyzed in this study was collected from 30 broiler chickens raised at the National Institute of Animal Science in Suwon, Korea. Chickens were supplied with crumbles *ad libitum* throughout the duration of the study, which comprised of 58.8% of corn, 28.0% soybean meal, 5.0% corn gluten meal, 1.1% wheat bran, 3.1% soybean oil, 1.4% calcium phosphate, 1.25% limestone, 0.5% vitamin mix, and traces of salt, lysine HCl, and DL-methionine. While antibiotics were not included, coccidiostat was added into the feed. This antiprotozoal agent prevents development of coccidiosis. Chicks received the set of vaccines typically used in the poultry industry – Marek's, Newcastle Disease and Infectious Bronchitis. Chickens were housed in pairs for the first 13 days to minimize stress and allow them to acclimatize to their environment in a single concrete floored pen bedded with sawdust and shavings, after which they were moved into individual, open-wire and temperature controlled cages. This was done to prevent competition for feed and minimize behavioral issues. Chickens were started on 22-23 hours per day of light from 1-day old, with this amount gradually decreasing for the duration of the study. Fecal samples analyzed in this study were aseptically collected from each chicken at two time points, days 1 and 35 of development, and stored in sterile bags at -80° C for further study. Both samples were used for analysis of the microbial community through pyrosequencing (Figure 3.1). For bacterial amplification, barcoded primers of 9F (5'-CCTATCCCCTGTGTGCCTTGGC AGTC-TCAG-AC-

AGAGTTTGATCMTGGCTCAG-3'; underlining sequence indicates the target region primer) and 541R (5'-CCATCTCATCCCTGCGT GTCTCCGAC-TCAG-X-AC-ATTACCGCGGCTGCTGG-3';'X'indicates the unique barcode for each subject) (http://oklbb.ezbiocloud.net/content/1001) were used. Amplification was carried out under the following conditions: initial denaturation at 95 ℃ for 5 min, followed by 30 cycles of denaturation at 95 ℃ for 30 sec, primer annealing at 55 ℃ for 30 sec, and extension at 72 ℃ for 30 sec, with a final elongation at 72 ℃ for 5 min. The PCR product was confirmed by gel electrophoresis on a 2% agarose gel and visualized under a Gel Doc system (BioRad, Hercules, CA, USA). The amplified products were purified with QIAquick PCR purification kit (Qiagen, Valencia, CA, USA). Equal concentrations of purified products were pooled together and short fragments (non-target products) were removed with Ampure beads kit (Agencourt Bioscience, MA, USA). The quality and product size were assessed on a Bioanalyzer 2100 (Agilent, Palo Alto, CA, USA) using a DNA 7500 chip. Mixed amplicons were conducted emulsion PCR, and then deposited on Picotiter plates. Sequencing was carried out at Chunlab, Inc. (Seoul, Korea) using the GS Junior Sequencing system (Roche, Branford, CT, USA) according to the manufacturer's instructions.

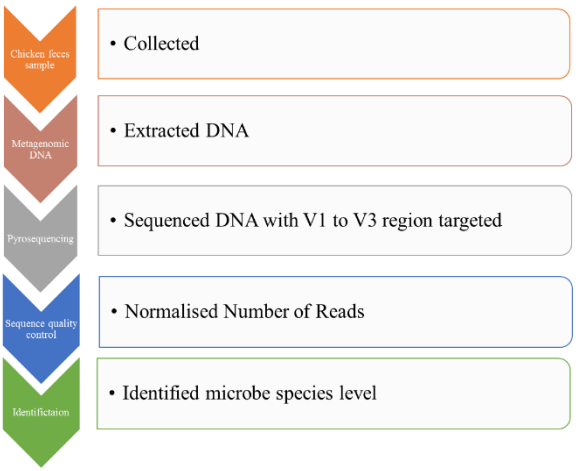


Figure 3.1. Scheme of bioinformatic pipeline for microbial population in the broiler chicken feces.

Bioinformatics analysis of pyrosequencing data

Basic analyses were conducted as previously described (Chun et al., 2010, Hur et al., 2011, Kim et al., 2012). The reads obtained from the different samples were sorted by the unique barcodes on each PCR product. The sequences of the barcode, linker, and primers were removed from the original sequencing reads. Any reads containing two or more ambiguous nucleotides, a low quality score (average score < 25), or reads shorter than 300bp were discarded. Potential chimera sequences were detected by the Bellerophone method, which compares the BLASTN search results between the forward half and reverse half of the sequences(Huber et al., 2004). After removing chimera sequences, the taxonomic classification of each read was assigned against the EzTaxon-e database (http://eztaxon-e.ezbiocloud.net)(Kim et al., 2012). This database contains 16S rRNA gene sequence of strains with valid published names and representative phylotypes of either cultured or uncultured entries in the GenBank database with complete hierarchical taxonomic classification from the phylum to the genus. The richness and diversity of samples was determined by operational taxonomic unit (OTUs), estimated OTU richness (ACE and Chao 1), diversity indices (Shannon and Simpson) and Goods Lib. Coverage at the 3% distance (Table 3.1). Random sub-sampling was conducted to equalize read size of samples to compare the different read sizes among samples. To compare OTUs between samples, shared OTUs were obtained using the XOR analysis of the CLcommunity software package (Chunlab Inc., Seoul, Korea).

Table 3.1 Number of validated sequences and comparison of phylotype coverage and diversity estimation at 97% similarity in operational taxonomic unit (OTUs), estimated OTU richness (ACE and Chao 1), diversity indices (Shannon and Simpson) and Goods Lib. Coverage in chicken fecal samples.

Sample	Valid reads	OTUs	Ace	Chao1	Shannon	Simpson	Goods Lib. Coverage
1day	14048	187	243.13	229.77	2.78	0.14	0.99
35days	4425	240	534.48	441.28	2.94	0.13	0.97

3.4 Result

Pyrosequencing data sets

A total of 26,049 sequence reads with valid barcodes were generated from the two samples obtained from 30 chickens at days 1 and 35 of development. The number of sequence reads was reduced to 22,792 after processing, and processed sequence reads were 484 bases long on average. The characteristics of both data sets are summarized in Table 3.2. Data collected showed that 99.87% of all processed reads from the 1-day old chickens were successfully assigned to the genus level. Similarly, 98.18% of all processed reads from 35-day old chickens were assigned to the genus level.

Table 3.2. Summary of pyrosequencing data

	1 day	35days	Total
Total number of reads	19854	6195	26049
Total number of selected reads	17187	5605	22792
Maximum sequence length	522	517	522
Minimum sequence length	140	298	140
Average sequence length	485	481	484

Microbial community structures in the feces of 1 and 35 days old chickens

The microbial composition of fecal samples at both time points is summarized at the phylum, class, order, and family level and portion of major phylogenetic types in Figure 3.2. Phyla in 1-day old chicken samples were represented by Firmicutes (68.61%), Proteobacteria (26.09%) and Streptophyta (5.30%), respectively. Class of *Bacilli* (54.05%), *Clostridia* (14.17%), and *Erysipelotrichi* (0.40%) constitute the Firmicutes. Class of *Gammaproteobacteria* was the most dominant Proteobacteria with an overall percentage of 26.08%. Among the 11 orders that were identified, *Lactobacillales* (54.05%), *Enterobacteriales* (26.06%), *Clostridiales* (14.16%) and *Poales* (5.17%) were the most abundant with percentage more than 1%. In the family level, total 18 groups were identified in this sample. Simultaneously, *Enterococcaceae* (51.64%), *Enterobacteriaceae* (26.06%), *Clostridiaceae* (14.91%), *Poacease* (5.17%) and *Lactobacillaceae* (2.52%) represented the main family phylotypes. Figure 3.2a shows that the main bacterial genera were *Enterobacter* (51.62%), *Escherichia* (26.02%), *Clostrodium* (14.08%) and *Lactobacillu*s (2.38%) among 30 kinds of genera identified. 6 phyla of Firmicutes, Proteobacteria, Actinobacteria, Tenericutes, Bacteroidetes and Cyanobacteria were identified in the feces sample from 35-days old chicken, among all of phyla, Firmicutes were highly overrepresented (99.05%). Similar to sample from 1-day old chickens, Class of *Bacilli* (72.65%), *Clostridia* (24.49%), and *Erysipelotrichi* (1.89%) were found to be major microorganisms, all of which belongs to phylum Firmicutes. In addition, the number of Class in

35 days were determined to be 13, which was larger than those in 1-day old samples. In order level, total 17 microorganisms were identified. Furthermore, main order groups consisted of *Lactobacillales* (72.14%), *Clostridiales* (24.47%) and *Turicibacter_O* (1.60%), respectively. Among the 40 family *phylotypes, Lactobacillaceae* (71.64%), *Peptostreptococcaceae* (15.46%), *Lachnospiraceae* (3.23%), *Ruminococcaceae* (3.05%), *Turicibacter* (1.60%) and *Arthromitus_F* (1.35%) was determined to be outstanding groups with representation of over 1%. As seen in Figure 3.3b, genera greatly increased in number compared to feces from 1-day old chickens (from 30 to 87). However, most of them belonged to *Lactobacillus* (71.48%) and *Clostrodium_g4* (15.38%). Also minor genus group with less than 1% were summarized in Appendix 6. Among 30 and 87 genera in each sample, only 4 genera showed the percentage over than 1% in both communities. However, its composition was different from each other. The change of microbiome at genus level was investigated through the comparison of the ratio between microorganisms with more than 1% portions (Figure 3.4). Overall, the major bacterial species were shown to be grouped together and also exhibit high levels of sequence similarity. In more detail, *Enterococcus*, *Escherichia* and *Clostridium* were determined to be dominant in the 1-day old group. On the other hand, *Lactobacillus*, and *Clostridium_g4* were the most prominent genus in 35-day old samples, belong to the Firmicutes group. This groups revealed obvious distinctions between two time points.

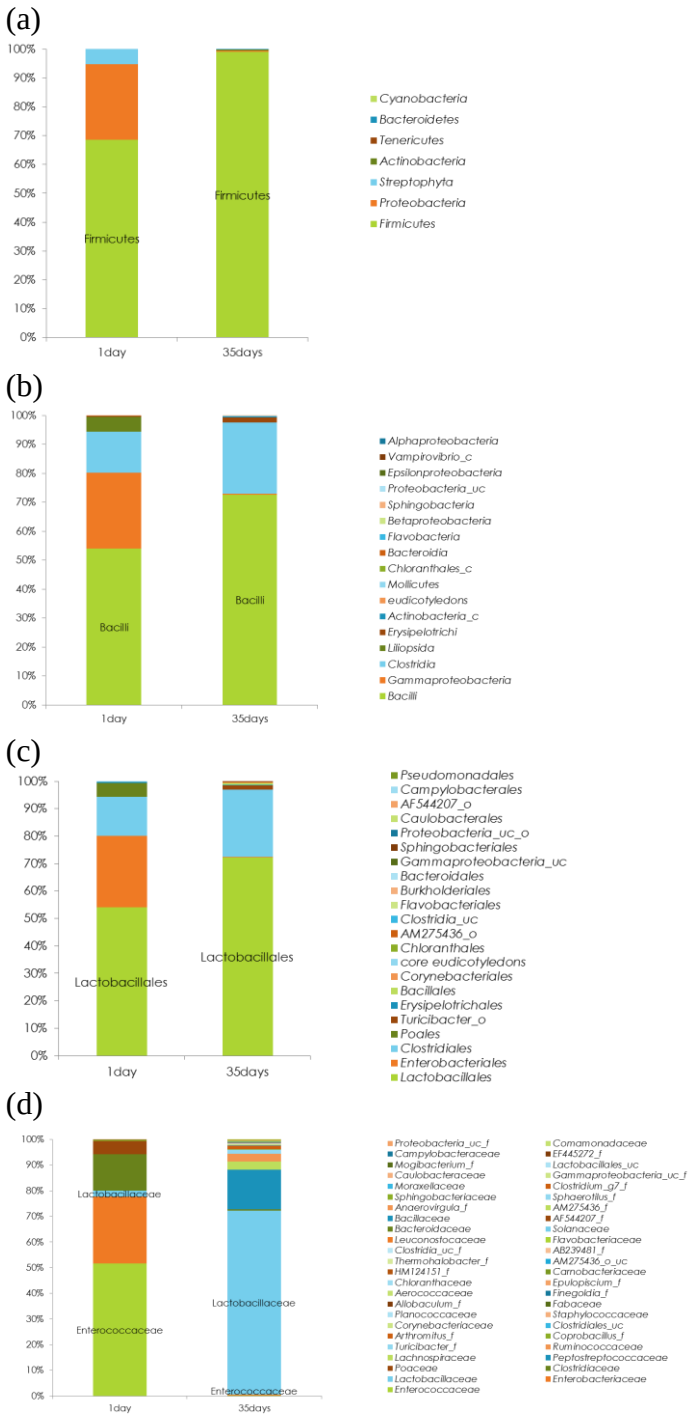
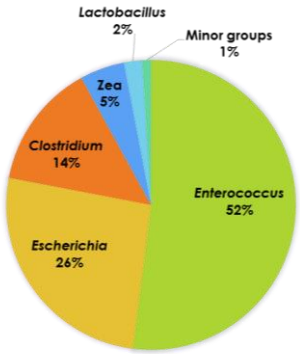
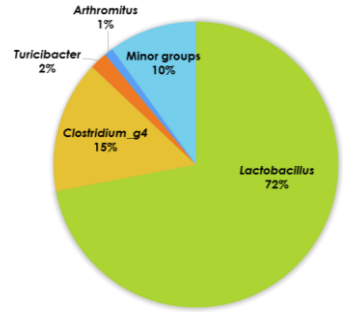


Figure 3.2. Phylogenetic distribution of each microbial community in chicken feces at the phylum (a), class (b), order (c) and family (d) level.



(a)



(b)

Figure 3.3. Major (>1%) groups of chicken fecal bacteria community structures at the genus level found in samples taken from 1-day old (a) and 35-day (b) old chickens.

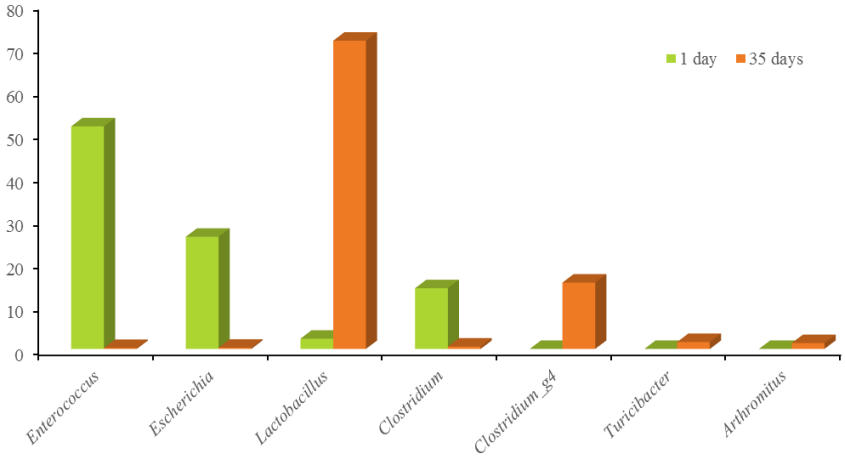


Figure 3.4. Comparison of dominant genus in both 1 and 35 days sample. Genus with 1% over percentage was selected as dominant groups.

3.5 Discussion

In this study, characterize important changes in fecal microbiota that occurred during the first 35 days of chicken development. A total of 9 and 13 classes, 11 and 17 orders, 18 and 40 families and 30 and 87 genera were identified on days 1 and 35, respectively. Oakley and coworkers analyzed the changes in the chicken cecal microbiome during 42 days growth and reported that the cecal bacterial community changed significantly and taxonomic richness and diversity at the genus level increased through growth time (Oakley et al., 2014). In a similar way, Danzeisen and colleagues examined the effects of the growth promoter virginiamycin and tylosine on the broiler chicken cecal microbiome and metagenome. Their estimation of diversity and richness control group without any additives clearly showed the tendency of increase in OTUs, Chao1, Shannon and Simpson index. The bacterial community analysis of chicken feces clearly showed proliferation of Firmicutes during growth. As expected, 1-day old chicken samples contained an excreted gut microbiome with a high percentage of Enterobacter and a limited percentage of Firmicutes such as Lactobacillus. However, as chickens reached maturity, the initial Firmicutes groups differentiated and became the dominant group as the chicken gut environment became anaerobic. Many studies had previously reported that Firmicutes is a major phylogenetic group in the gut flora of various animals such as duck, pig, and chicken (Guo et al., 2008, Danzeisen et al., 2011, Eeckhaut et al., 2011, Becker et al., 2014). Results have revealed similarities in

family, genus and species from feces and gastrointestinal tract samples. For instance, high abundance of orders *Clostridiales, Lactobacillales, Bacteroidales, Bifidobacteriales, Enterobacteriales, Erysipelotrichales, Coriobacteriales, Desulfovibrionales, Burkholderiales, Campylobacterales* and *Actinomycetales* was detected in the feces of hens(Videnska et al., 2013). *Enterobacteriales, Lactobacillales* and *Pseudomonadales* can be found in the gut of healthy children (Gupta et al., 2011). Pyrosequencing analysis of our samples was able to identify the same dominant species detected using clone libraries (Orcutt et al., 2009). A total of 30 genera were identified in 1-day old chicken fecal samples. This complex microbial community was biased towards a few dominant genera such as *Enterobacter* (51.62%), *Escherichia* (26.02%), *Clostrodium* (14.08%) and *Lactobacillu*s (2.38%). The bacterial fecal community of 35-day old chickens consisted of 87 different gerera. However, this population was also restricted to a few major genera that included *Lactobacillus* (71.48%) and *Clostrodium_g4* (15.38%). The global composition of fecal microbiota was similar to that reported in previous studies of chicken, with Firmicutes dominating the makeup, followed by Bacteroidetes and Proteobacteria (Eeckhaut et al., 2011, Becker et al., 2014). In genus level, some pathogens such as *Clostridium, Enterococcus* and *Escherichia* were found in the 1-day old sample in high titer. With the exception of *Clostridium* and *Escherichia* the same pathogens were found in the 35-day old sample group but with low titer. An overwhelming percentage of Firmicutes (99.05%) was identified in fecal samples collected from 35-day old chickens. This was a

surprising result, given that previous studies found that Firmicutes gradually decreased with development in monensin/virginiamycin treated groups (Danzeisen et al., 2011). Probiotics and antiprotozoal agents are routinely introduced to the digestive tract of poultry through feed in order to prevent development of disease(Kabir, 2009). In fact, given concerns over side-effects of the use of antibiotics, recent years have seen a preference for probiotics in the poultry industry (Trafalska & Grzybowska, 2003, Griggs & Jacob, 2005, Nava et al., 2005). Chickens typically receive complete gut flora from their mother's feces, which protects them from infection. However, broiler chickens are typically raised in sterile incubators and deprived of contact with their mothers and other adults. Supplying probiotics immediately after birth is especially important, as these features of the poultry production process makes their protective gut microflora particularly susceptible to change and damage (Fuller, 2001). While a diet supplemented with coccidiostat such as that of poultry in this study is typical, it is important to consider that administration of antiprotozoal drugs may be an important contributing factor in the specific changes of microbiota reported. However, although probiotic and antibiotic administration does not provide a true depiction of developmental changes that would occur naturally, following the standard protocol used in the poultry production process allows this investigation of microbial ecology to be applied to the poultry industry. Pyrosequencing analyses of the excreted microbiome of 1-day old and 35-days old chickens identified important changes in microbiota composition. Previous studies revealed a dynamic shift of the community

structure of the chicken gut depending on the hatching and rearing environment (Sekelja et al., 2012). Results of the present study confirm this bacterial community shift. Interestingly, no common phylogenetic groups were found between samples from each time point except for *Lactobacillus*. Results reveal dramatic microbial changes in composition during a relatively short period of broiler chicken development and that these changes are indeed highly dynamic. It anticipate that further understanding of the developmental dynamics of fecal microbiota during chicken development, impacted by common probiotics, can help in efforts to eliminate contamination of poultry through prevention of intestinal infections.

This chapter is published in Asian-Australasian Journal of Animal Sciences
as a partial fulfillment of Sooyeon Lim's Ph.D program.

Chapter 4. Metagenome analysis of protein domain collocation within cellulose genes of goat rumen microbes

4.1 Abstract

Goats have unique habits, which include feeding on unconventional tree leaves. Thus, goats are expected to host distinct bacterial communities with cellulose-degrading enzyme activity in their rumen. In this study, protein domains with cellulase activity in goat rumen microbes were investigated using metagenomic and bioinformatic analyses. After the complete genome of the goat rumen microbe was obtained using a shotgun sequencing method, 217,892,109 pair reads were filtered, including only those with 70% identity, 100-bp matches, and thresholds below E−10 using METAIDBA. These filtered contigs were assembled and annotated using blastN against the NCBI nucleotide database. As a result, a microbial community structure with 439 genera was analyzed, among which *Prevotella* and *Butyrivibrio* were the dominant groups. In parallel, 201 sequences related with cellulase activities (EC.3.2.1.4) were obtained through blast searches using the enzyme.dat file provided by the NCBI database. After translating the nucleotide sequence into a protein sequence using Interproscan, 28 protein domains with cellulase activity were identified using the HMMER package with threshold E values below 10–5. Cellulase activity protein domain profiling showed that the major protein domains such as lipase GDSL, celluase, and Glyco hydro 10 were present in bacteria with strong cellulase activities. Furthermore, correlation plot clearly displayed the strong positive correlation between some protein domain groups, which was indicative of microbial adaption in the goat rumen based on

feeding habits. This is the first metagenomic analysis of cellulase activity protein domains using bioinformatics from the goat rumen.

4.2 Introduction

Goats have an extremely varied diet including the tips of woody shrubs, trees, and lignocellulosic agricultural by-products. Symbiont microbes in the rumen of these herbivores play key roles in providing the hosts with various nutrients. Enzymes secreted by rumen microbes are essential for the conversion of cellulose and hemi-cellulose into simple sugars, which are metabolized to volatile fatty acids by rumen microbes. Produced volatile fatty acids serve as energy sources for ruminants. Many studies have investigated the symbiotic microorganisms in the rumen because of their link to economically or environmentally important traits such as feed conversion efficiency, methane production (Hegarty, 1999, Guan et al., 2008, Hess et al., 2011). A key challenge in this study was identifying rumen microbial profiles, which are associated and potentially predictive of these traits. Thus, methods for profiling the rumen microbial population should be relatively inexpensive and efficient to allow a large number of individuals to be profiled (Ross et al., 2012). Untargeted rumen bacterial communities contain numerous novel gene sequences based on deep sequencing of pooled samples of true biological variation. The rumen metagenome profile included the counts of reads that aligned to each contig, which could be analyzed using metagenomic tools and correlation plots. The composition of the microbial population differs between goat species and based on their diet. Analysis of microorganisms in the rumen fluid of different herbivores revealed bacteria (1010–1011 cells/ml,

representing more than 50 genera), ciliate protozoa (104–106/ml, from 25 genera), anaerobic fungi (103–105 zoospores/ml, representing six genera), and bacteriophages (108–109/ml). These numbers represented only a small fraction of the microbial species in rumens of animals on fiber-based diets since less than 10–20% of microbial populations are cultivable on synthetic media(Zhou et al., 1996). However, metagenomic research has generated genetic information on the entire microbial community, which is important because 99% of microbes cannot be isolated or cultured. The metagenomic method provides a global microbial gene pool without the need to culture of the microorganisms. In this study, analyzed the complete genome of goat rumen microbes obtained using a shotgun sequencing method. This differed from previous studies on microbes based on 16 rRNA. Also, our results were filtered under strict conditions and provided high-quality results on the rumen microbe community and cellulose activity protein domains.

4.3 Materials and methods

Sampling and extraction of genomic DNA

Rumen fluid was collected from a 1-year-old Korean native goat and Saanen hybrid raised on Timothy (Phleumpratense) hay at a private goat farm in the Cheonan City area and slaughtered at a local slaughter house. Rumen fluid was filtered through four layers of cheesecloth. Genomic DNA was isolated from rumen fluid using the Wizard Genomic DNA Purification Kit (Promega, U.S.) according manufacturer's protocol. Gel electrophoresis was performed with 1% agarose gel at 50V for 2 hours to check both quality and quantity of isolated genomic DNA.

DNA shotgun paired-end library preparation

Random DNA fragmentation was performed using the Covaris S2 System, and the DNA library was prepared using TruSeqDNA Sample Prep.Kit (Illumina, U.S). Briefly, DNA fragments were repaired to blunt-ended DNA by fill-in and exonuclease after A-tailing was conducted to prevent the formation of adapters, dimers, and concatemers. Adaptors were ligated to genomic DNA inserts at a molar ratio of 10:1. The DNA samples were then amplified via polymerase chain reaction (PCR) using two universal primers. One primer contained an attachment site for the flow cell and the other contained sequencing sites for the index read. After gel electrophoresis of the PCR

product, 600–700-bp fragments (including the insert and adapter) were selected and purified for genomic sequencing.

Genomic sequencing

Genomic DNA sequences were generated using the Illumina Hiseq2000 platform. Briefly, only library fragments with proper adapters at both ends were amplified using P5 and P7 primers on the flow cell. Clonal clusters were generated using TreSeq PE Cluster kitV3-cBot-HS (ILPE-401-3001; Illumina). Using the HiSeq2000 platform with TruSeq SBS Kit v3-HS (200 cycles; ILFC-401-3001; Illumina) 435,784, 218 reads were obtained.

Metagenomic bioinformatics application

Each pair read, scaffold, and contig of the shotgun sequencing of goat rumen microbes was summarized in Figure 4.1.and Table 4.1. Whole genomic DNA of collected goat rumen microbes were extracted for Illumina sequencing without DNA targeting. This shotgun sequencing generated 217,892,109 pair reads, which were filtered based on 70% identity, more than a 100-bp match, and a threshold below E–10 based on METAIDBA. These filtered 1,373,011 scaffolds were assembled and annotated to 114,031 contigs using blastN against the NCBI nucleotide database. Subsequent D/B matching with unique ID of each sequence generated 5,411 contigs. Then, same ID name was sorted, which confirmed the 1,431 contigs. Finally, these annotated genomic sequences were

assigned for both identification of microbial community and cellulose-like protein domains

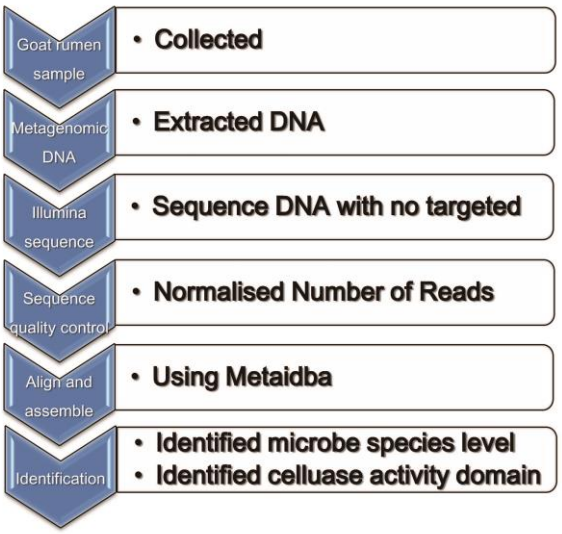


Figure 4.1. Each pair read, scaffold and contig from the shotgun sequencing of goat rumen microbes

Table 4.1 Assembly and annotation statistics

Assembly	Annotation
Total pair reads	217,892,109
Scaffolds	1,373,011
Contigs	114,031

4.4 Results

Microbial community structure in goat rumen

The isolated genes in rumen fluid were classified into a total of 1,704 organisms, among which each 181 and 1431 ID corresponded to plant and bacteria, respectively. Using the METAIDBA metagenomic bioinformatic program, 114,031 sequences were classified into 439 genera; their population structure at the genus level is graphically depicted in Figure 4.2. *Prevotella* and *Butyrivibrio* bacteria were the dominant populations, accounting for 18% and 14%, respectively. The majority of goat rumen bacteria identified in this study have been previously reported in the rumens of cow or lamb, such as *Prevotella ruminicola 23, Butyrivibrio proteoclasticus B316, and Butyrivibrio fibrisolvens* (Bryant & Small, 1956, van Gylswyk & vander Toorn, 1986, McKain et al., 1992, Moon et al., 2008). Also, some microorganisms such as *butyrate-producing bacterium SS3/4* have been identified in the human colon. Previous studies have revealed the detailed rumen metabolism of *Fibrobacter succinogenes subsp. succinogenes S85 and Selenomonas ruminantium* (Heinrichova et al., 1989, Chow & Russell, 1992).

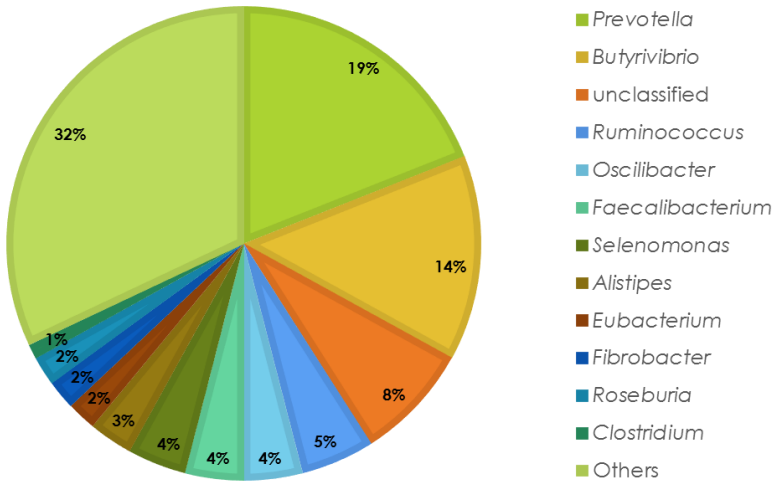


Figure 4.2. Analysis of the goat rumen microbial community structure at the genus level.

Profile of protein domains with cellulase activity

Cellulase protein ID was obtained from the enzyme.dat file provided by the NCBI database. As a result, 201 sequences related with cellulase activity were obtained through blast searches using the NCBI BLAST program. After translating the nucleotide sequence into a protein sequence using Interproscan, 28 protein domains with cellulase activity were identified using the HMMER package with threshold E values below 10–5. These 28 protein domains with cellulase activity are summarized in Table 4.2. After 28 protein domains with cellulase activity were identified, the richness of each domain was analyzed (Figure 4.3). Some of protein domains were overlapped to same part of sequences and also counted. The dominant bacteria had a larger number of protein domains, which suggested that strong cellulase activities were related to bacterial survival in the goat rumen. Protein domains with high richness such as lipase GDSL, cellulase, and Glyco hydro 10 were also identified in the goat rumen microbes. Next, the number of protein domains in each microbe was investigated (Figure 4.4). Prevalent bacteria such as *Prevotella* and *Butyrivibrio* contained a large number of cellulase protein domains, implying that these bacteria play a role in the degradation of cellulose in the goat rumen. Finally, the protein domain ratio in each microbial group was analyzed (Figure 4.5). The dominant bacteria showed a ratio greater than 1, suggesting that they have high cellulose activity. A correlation plot among 28 protein domains (Figure 4.6) confirmed the strong positive correlation between some protein domain groups. For example, CHB_HEX_c and CHB_HEX_c-1, CHB_HEX_c and

fn3 asso, and CHB_HEX_c -1 and fn3_asso had a positive correlation greater than 0.99. Another group of lipase GDSLs, lipase GDSL_2, also showed a positive correlation greater than 0.99. To determine whether the goat rumen microbe profile was predictive of the rumen fluid metagenome profile, it correlated every rumen metagenome profile with every cellulose activity protein domain. Determined whether the correlations were higher for samples from the same animal than for between animal samples. The results suggested that rumen fluid samples had strong correlations with each protein domain. Microbial community structure and specific protein domains with cellulase activity in the goat rumen have been identified using metagenomic analysis with both shotgun sequencing and bioinformatics. This study demonstrated that specific dominant bacterial taxonomy and protein domains have strong positive correlations, suggesting adaption to the unique feeding habits of goats.

Table 4.2. List of protein domains with cellulase activity in goats.

Domain	Accession	References
CBM_11	Pfam03425	no ref
CBM_2	Pfam 00553	Xu, et al., 1995 (Xu et al., 1995)
CBM_3	Pfam00942	Poole, et al., 1992, Tormo, et al., 1996 (Poole et al., 1992, Tormo et al., 1996)
CBM_4_9	Pfam02018	Johnson, et al., 1996 (Johnson et al., 1996)
CBM49	Pfam09478	Mosbah, et al., 2000 (Mosbah et al., 2000)
CBM_5_12	Pfam02839	no ref
CBM_X2	Pfam03442	Mosbah, et al., 2000, Kosugi, et al., 2004 (Mosbah et al., 2000, Kosugi et al., 2004)
elD_N	Pfam02927	Dominguez, et al., 1996 (Dominguez et al., 1996)
Cellulase	Pfam00150	no ref
Cellulase-like	Pfam12876	no ref
CHB_HEX_C	Pfam03174	Tews, et al., 1996 (Tews et al., 1996)
CHB_HEX_C_1	Pfam13290	no ref
CIA30	Pfam08547	Janssen, et al., 2002 ,Walker, et al., 1992 (Walker et al., 1992, Janssen et al., 2002)
Dockerin_1	Pfam00404	Lytle, et al., 2000 , Shoham, et al., 1999 (Shoham et al., 1999, Lytle et al., 2000)
DPBB_1	Pfam 03330	Takase, et al., 1987 , Castillo, et al., 1999 , Mizuguchi, et al., 1999 (Takase et al., 1987, Castillo et al., 1999, Mizuguchi et al., 1999)
fn3	Pfam00041	Bazan, et al., 1990, Kornblihtt, et al., 1985, Little, et al., 1994(Kornblihtt et al., 1985, Bazan, 1990, Little et al., 1994)
Fn3_assoc	Pfam13287	no ref
Glyco_hydro_10	Pfam00331	no ref
Glyco_hydro_26	Pfam02156	no ref
Glyco_hydro_44	Pfam12891	Kitago, et al., 2007(Kitago et al., 2007)
Glyco_hydro_45	Pfam02015	no ref
Glyco_hydro_48	Pfam02011	no ref
Glyco_hydro_8	Pfam01270	Alzari, et al., 1996(Alzari et al., 1996)
Glyco_hydro_9	Pfam00759	no ref
I-set	Pfam07679	no ref
Lipase_GDSL	Pfam 00657	Brick, et al., 1995(Brick et al., 1995)
Lipase_GDSL_2	pfam13472	Molgaard, et al., 2000(Molgaard et al., 2000)
SLH	Pfam00395	Mesnage, et al., 2000(Mesnage et al., 2000)

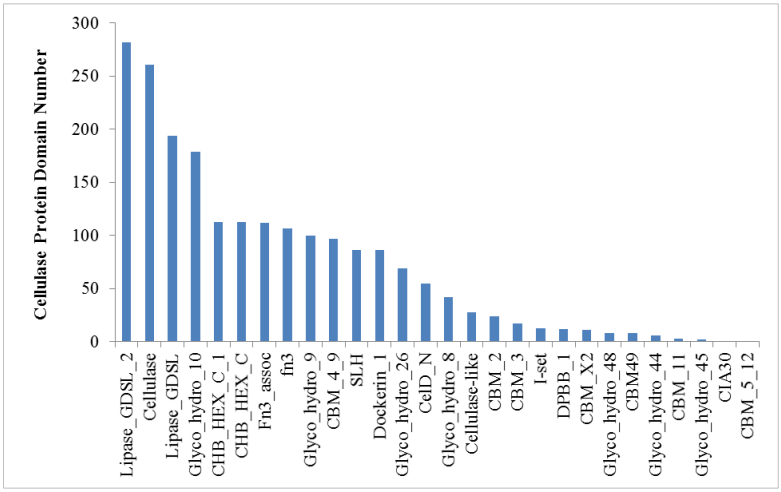


Figure 4.3. Richness of each protein domain with cellulase activity in the goat rumen microbial population.

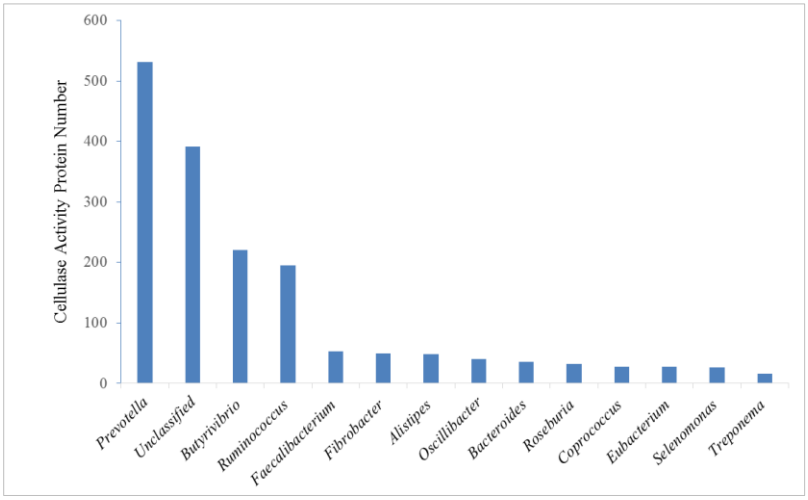


Figure 4.4. Protein domains with cellulase activity were present in over 1% of the dominant microbes

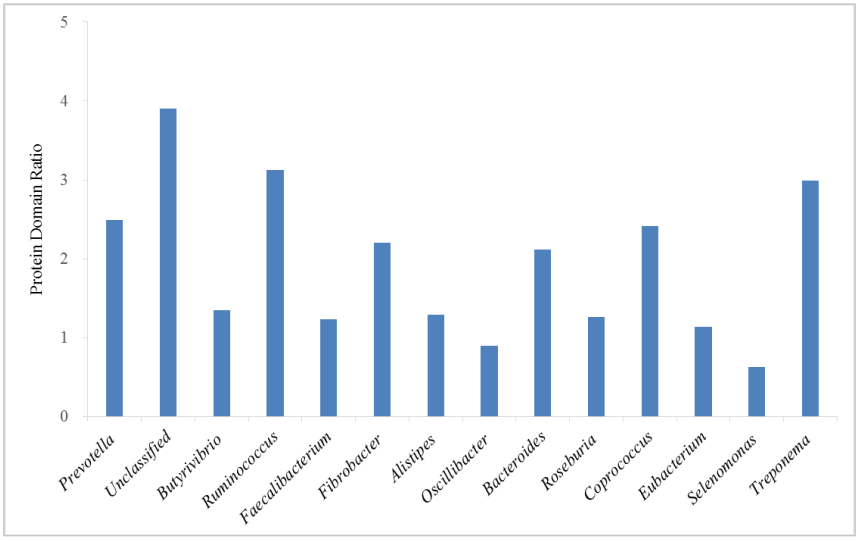


Figure 4.5. Ratio of protein domain in each bacteria at the genus level.

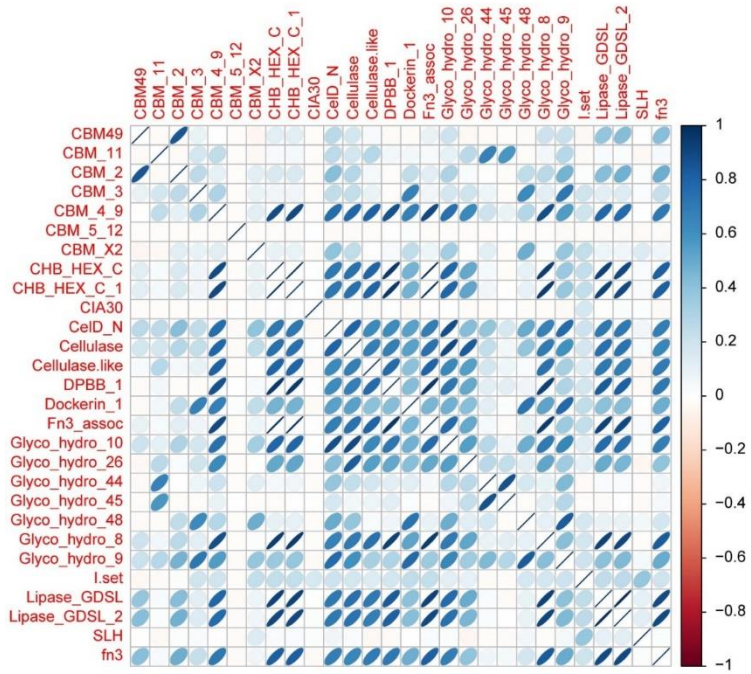


Figure 4.6. Correlation plot between each cellulase activity protein domain in goat rumen.

4.5 Discussion

In this study, microbial community structure and specific protein domains with cellulase activity in the goat rumen were identified using metagenomic analysis with both shotgun sequencing and bioinformatics. As a result, the presence of both specific dominant bacterial genus such as *Prevotella, Butyrivibrio* were identified in goat rumen fluid. At the same time, 28 protein domains with cellulose-like activity such as lipase GDSL, cellulase, and Glyco hydro 10 were identified with strong positive correlations, suggesting adaption to the unique feeding habits of goats. Cunha and coworkers reported that bacteria from the phyla of Bacteroidetes and Firmicutes were predominant in the Moxotó goat rumen. Furthermore, the overall dominant classes in the rumen were identified to be *Clostridia* (37.9%) and *Bacteroidia* (56.3%), which are known to play a role in plant fiber degradation in other ruminants. Among 12 major genus which was identified in this study, *Prevotella* (18 %) and *Alistipes* (3%) represented at the phylum of Bacteroidetes. Genus of *Butyrivibrio* (14%), *Ruminococcus* (5%), *Oscillibacter* (4 %), *Faecalibacterium* (4 %), *Selenomonas* (4 %), *Eubacterium* (2 %), *Roseburia* (1 %), and *Clostridium* (1 %) belonged to the Firmicutes phylum. At the class level comparison, *Butyrivibrio, Ruminococcus, Oscillibacter, Faecalibacterium, Eubacterium, Roseburia*, and *Clostridium* constitute of *Clostridia*. *Prevotella* and *Alistipes* were determined to the class of *Bacteroidia*. Despite of some difference between detailed community profiles, the overall tendency in the goat rumen microbiome such

as major groups could be said to be similar with those of other study. Degradation of cellulose in the fibulous matrix is one of the major research topics in the rumen animals such as sheep, cow and goat. For other ruminant animals, Toyoda and coworkers analyzed the cellulose-binding proteins from sheep rumens, which consisted of endo-glucanases, proteins from fiber degrading bacterium and exo-glucanases, respectively (Toyoda et al., 2009). For cattle, constructed metagenomic library and identified 22 clones with distinct hydroylic activities such as 12 esterases, nine endo-β-1,4-glucanases and one cyclodextrin (Ferrer et al., 2005). Considering the close correlation between rumen microbial ecology and its enzymatic functions according to the other ruminal livestock (Krause et al., 2013), list of cellulase-like protein domain list of this study can provide a clue to the characterization of Korean native goat rumen.

General Discussion

By using next generation sequencing and metagenomic analysis, correlation between intestinal microbiome with environmental factors of commercially important animals could be investigated in perspective of phylogeny and metabolism. This study applied this research frame to the bovine, chicken and goat, respectively.

For bovine, sequencing the genomes of rumen microbes, determining the role of the genes and identifying the potential applications are the great deal for researchers to understand the microbiology of the rumen. Rumen solid and rumen liquid of cattle rumen was analyzed using high-throughput pyrosequencing of 16S rRNA gene PCR amplicons and a subsequent bioinformatics pipeline.

For chicken, quantitatively analyze the fecal microbial community using next generation sequencing (NGS) techniques in this study. Fecal samples were collected from 30 broiler chickens at two time points- days 1 and 35 of development. 454 pyrosequencing was conducted on 16S rRNA extracted from each sample, and microbial population dynamics were investigated using various automated bioinformatics pipelines. From this study, Bacteroidetes, Firmicutes and Fibrobacteria were present in higher concentrations in rumen solid than in rumen liquid, indicating their major role in the degradation of plant fiber.

For goat, protein domains with cellulase activity in goat rumen microbes were investigated using metagenomic and bioinformatic analyses. 201 sequences related with cellulase activities (EC.3.2.1.4) were obtained. In parallel, 28 protein domains with cellulase activity were identified. Based on these result, correlation plot showed the strong positive correlation between cellulose activity protein domain groups.

These analyses can contribute to extend our understanding about microbial community in animal digestion system, which is one of the important issues in the livestock industry. Thus, synergy between NGS and Metagenomic could produce fundamental new information for animal science at present and in future.

Reference

Ahmadian A, Ehn M & Hober S (2006) Pyrosequencing: history, biochemistry and future. *Clinica chimica acta* **363**: 83-94.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Alzari PM, ne Souchon H & Dominguez R (1996) The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from Clostridium thermocellum. *Structure* **4**: 265-275.

Backhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF & Gordon JI (2004) The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A* **101**: 15718-15723.

Bazan JF (1990) Structural design and molecular evolution of a cytokine receptor superfamily. *Proc Natl Acad Sci U S A* **87**: 6934-6938.

Becker AA, Hesta M, Hollants J, Janssens GP & Huys G (2014) Phylogenetic analysis of faecal microbiota from captive cheetahs reveals underrepresentation of Bacteroidetes and Bifidobacteriaceae. *BMC Microbiol* **14**: 43.

Bhasin M, Garg A & Raghava GP (2005) PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinformatics* **21**: 2522-2524.

Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC & Hugenholtz P (2007) CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**: 209.

Brick DJ, Brumlik MJ, Buckley JT, Cao JX, Davies PC, Misra S, Tranbarger TJ & Upton C (1995) A new family of lipolytic plant enzymes with members in rice, arabidopsis and maize. *FEBS Lett* **377**: 475-480.

Brulc JM, Antonopoulos DA, Miller MEB, Wilson MK, Yannarell AC, Dinsdale EA, Edwards RE, Frank ED, Emerson JB & Wacklin P (2009) Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc Natl Acad Sci U S A* **106**: 1948-1953.

Bryant MP & Small N (1956) The anaerobic monotrichous butyric acid-producing curved rod-shaped bacteria of the rumen. *J Bacteriol* **72**: 16-21.

Castillo RM, Mizuguchi K, Dhanaraj V, Albert A, Blundell TL & Murzin AG (1999) A six-stranded double-psi beta barrel is shared by several protein superfamilies. *Structure* **7**: 227-236.

Chow JM & Russell JB (1992) Effect of pH and Monensin on Glucose Transport by Fibrobacter succinogenes, a Cellulolytic Ruminal Bacterium. *Appl Environ Microbiol* **58**: 1115-1120.

Chun J, Kim KY, Lee JH & Choi Y (2010) The analysis of oral microbial communities of wild-type and toll-like receptor 2-deficient mice using a 454 GS FLX Titanium pyrosequencer. *BMC Microbiol* **10**: 101.

Chun J, Lee JH, Jung Y, Kim M, Kim S, Kim BK & Lim YW (2007) EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int J Syst Evol Microbiol* **57**: 2259-2261.

Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan P-L, Briese T, Hornig M & Geiser DM (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**: 283-287.

Cummings JH & Macfarlane GT (1997) Role of intestinal bacteria in nutrient metabolism. *JPEN J Parenter Enteral Nutr* **21**: 357-365.

Daniel R (2005) The metagenomics of soil. *Nat Rev Microbiol* **3**: 470-478.

Danzeisen JL, Kim HB, Isaacson RE, Tu ZJ & Johnson TJ (2011) Modulations of the chicken cecal microbiome and metagenome in response to anticoccidial and growth promoter treatment. *PLoS One* **6**: e27949.

Degnan PH & Ochman H (2012) Illumina-based analysis of microbial community diversity. *Isme j* **6**: 183-194.

Dela Torre R, Larkin J, Singer A & Meller A (2012) Fabrication and characterization of solid-state nanopore arrays for high-throughput DNA sequencing. *Nanotechnology* **23**: 385308.

Dominguez R, Souchon H, Lascombe M & Alzari PM (1996) The crystal structure of a family 5 endoglucanase mutant in complexed and uncomplexed forms reveals an induced fit activation mechanism. *J Mol Biol* **257**: 1042-1051.

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE & Relman DA (2005) Diversity of the human intestinal microbial flora. *Science* **308**: 1635-1638.

Eeckhaut V, Van Immerseel F, Croubels S, De Baere S, Haesebrouck F, Ducatelle R, Louis P & Vandamme P (2011) Butyrate production in phylogenetically diverse Firmicutes isolated from the chicken caecum. *Microbial biotechnology* **4**: 503-512.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P & Bettman B (2009) Real-time DNA sequencing from single polymerase molecules. *Science* **323**: 133-138.

Emanuelsson O, Brunak S, von Heijne G & Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**: 953-971.

Ferrer M, Golyshina OV, Chernikova TN, *et al.* (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol* **7**: 1996-2010.

Forsberg CW, Cheng K-J & White BA (1997) Polysaccharide degradation in the rumen and large intestine. *Gastrointestinal microbiology,* 319-379. Springer.

Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N & Pace NR (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* **104**: 13780-13785.

Fuller R (2001) The chicken gut microflora and probiotic supplements. *The Journal of Poultry Science* **38**: 189-196.

Glass EM, Wilkening J, Wilke A, Antonopoulos D & Meyer F (2010) Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc* **2010**: pdb prot5368.

Griggs J & Jacob J (2005) Alternatives to antibiotics for organic poultry production. *The Journal of Applied Poultry Research* **14**: 750-756.

Guan LL, Nkrumah JD, Basarab JA & Moore SS (2008) Linkage of microbial ecology to phenotype: correlation of rumen microbial ecology to cattle's feed efficiency. *FEMS Microbiol Lett* **288**: 85-91.

Guo X, Xia X, Tang R, Zhou J, Zhao H & Wang K (2008) Development of a real-time PCR method for Firmicutes and Bacteroidetes in faeces and its application to quantify intestinal population of obese and lean pigs. *Letters in Applied Microbiology* **47**: 367-373.

Gupta SS, Mohammed MH, Ghosh TS, Kanungo S, Nair GB & Mande SS (2011) Metagenome of the gut of a malnourished child. *Gut Pathog* **3**: 7.

Gužvić M (2013) The History of DNA Sequencing/ISTORIJAT SEKVENCIRANJA DNK. *Journal of Medical Biochemistry* **32**: 301-312.

Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, DiMeo J & Efcavitch JW (2008) Single-molecule DNA sequencing of a viral genome. *Science* **320**: 106-109.

Hegarty RS (1999) Reducing rumen methane emissions through elimination of rumen protozoa. *Australian Journal of Agricultural Research* **50**: 1321-1328.

Heinrichova K, Wojciechowicz M & Ziolecki A (1989) The pectinolytic enzyme of Selenomonas ruminantium. *J Appl Bacteriol* **66**: 169-174.

Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F & Zhang T (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**: 463-467.

Hess M, Sczyrba A, Egan R, *et al.* (2011) Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* **331**: 463-467.

Hijova E & Chmelarova A (2007) Short chain fatty acids and colonic health. *Bratisl Lek Listy* **108**: 354-358.

Holmes I, Harris K & Quince C (2012) Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* **7**: e30126.

Hooper LV & Gordon JI (2001) Commensal host-bacterial relationships in the gut. *Science* **292**: 1115-1118.

Huber T, Faulkner G & Hugenholtz P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**: 2317-2319.

Hugenholtz P (2002) Exploring prokaryotic diversity in the genomic era. *Genome Biol* **3**: Reviews0003.

Hugenholtz P & Tyson GW (2008) Microbiology: metagenomics. *Nature* **455**: 481-483.

Hur M, Kim Y, Song HR, Kim JM, Choi YI & Yi H (2011) Effect of genetically modified poplars on soil microbial communities during the phytoremediation of waste mine tailings. *Appl Environ Microbiol* **77**: 7611-7619.

Huws SA, Kim EJ, Lee MR, Scott MB, Tweed JK, Pinloche E, Wallace RJ & Scollan ND (2011) As yet uncultured bacteria phylogenetically classified as Prevotella, Lachnospiraceae incertae sedis and unclassified Bacteroidales, Clostridiales and Ruminococcaceae may play a predominant role in ruminal biohydrogenation. *Environ Microbiol* **13**: 1500-1512.

Jakhesara S, Koringa P, Ramani U, Ahir V, Tripathi A, Soni P, Singh K, Bhatt V, Patel J & Patel M (2010) Comparative study of tannin challenged rumen

microbiome in goat using high throughput sequencing technology. *Dev Microbiol Mol Biol* **1**: 95-106.

Janssen R, Smeitink J, Smeets R & van Den Heuvel L (2002) CIA30 complex I assembly factor: a candidate for human complex I deficiency? *Hum Genet* **110**: 264-270.

Johnson PE, Joshi MD, Tomme P, Kilburn DG & McIntosh LP (1996) Structure of the N-terminal cellulose-binding domain of Cellulomonas fimi CenC determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **35**: 14381-14394.

Jones BV, Begley M, Hill C, Gahan CG & Marchesi JR (2008) Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. *Proc Natl Acad Sci U S A* **105**: 13580-13585.

Kabir S (2009) The role of probiotics in the poultry industry. *International journal of molecular sciences* **10**: 3531-3546.

Kang SH, Evans P, Morrison M & McSweeney C (2013) Identification of metabolically active proteobacterial and archaeal communities in the rumen by DNA- and RNA-derived 16S rRNA gene. *J Appl Microbiol* **115**: 644-653.

Kim BS, Kim JN, Yoon SH, Chun J & Cerniglia CE (2012) Impact of enrofloxacin on the human intestinal microbiota revealed by comparative molecular analysis. *Anaerobe* **18**: 310-320.

Kim OS, Cho YJ, Lee K, *et al.* (2012) Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int J Syst Evol Microbiol* **62**: 716-721.

Kitago Y, Karita S, Watanabe N, Kamiya M, Aizawa T, Sakka K & Tanaka I (2007) Crystal structure of Cel44A, a glycoside hydrolase family 44 endoglucanase from Clostridium thermocellum. *J Biol Chem* **282**: 35703-35711.

Koike S, Yoshitani S, Kobayashi Y & Tanaka K (2003) Phylogenetic analysis of fiber-associated rumen bacterial community and PCR detection of uncultured bacteria. *FEMS Microbiol Lett* **229**: 23-30.

Kornblihtt AR, Umezawa K, Vibe-Pedersen K & Baralle FE (1985) Primary structure of human fibronectin: differential splicing may generate at least 10 polypeptides from a single gene. *Embo j* **4**: 1755-1759.

Kosugi A, Amano Y, Murashima K & Doi RH (2004) Hydrophilic domains of scaffolding protein CbpA promote glycosyl hydrolase activity and localization of cellulosomes to the cell surface of Clostridium cellulovorans. *J Bacteriol* **186**: 6351-6359.

Kotera M, Hirakawa M, Tokimatsu T, Goto S & Kanehisa M (2012) The KEGG databases and tools facilitating omics analysis: latest developments involving human diseases and pharmaceuticals. *Methods Mol Biol* **802**: 19-39.

Kotera M, Hirakawa M, Tokimatsu T, Goto S & Kanehisa M (2012) The KEGG databases and tools facilitating omics analysis: latest developments involving human diseases and pharmaceuticals. *Next Generation Microarray Bioinformatics,* 19-39. Springer.

Krause DO, Nagaraja TG, Wright AD & Callaway TR (2013) Board-invited review: Rumen microbiology: leading the way in microbial ecology. *J Anim Sci* **91**: 331-341.

Krogh A, Larsson B, von Heijne G & Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567-580.

Kudva IT, Stanton TB & Lippolis JD (2014) The Escherichia coli O157:H7 bovine rumen fluid proteome reflects adaptive bacterial responses. *BMC Microbiol* **14**: 48.

Leschine SB (1995) Cellulose degradation in anaerobic environments. *Annu Rev Microbiol* **49**: 399-426.

Li KB (2003) ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics* **19**: 1585-1586.

Liakopoulos TD, Pasquier C & Hamodrakas SJ (2001) A novel tool for the prediction of transmembrane protein topology based on a statistical analysis of the SwissProt database: the OrienTM algorithm. *Protein Eng* **14**: 387-390.

Lim S, Seo J, Choi H, Yoon D, Nam J, Kim H, Cho S & Chang J (2013) Metagenome Analysis of Protein Domain Collocation within Cellulase Genes of Goat Rumen Microbes. *Asian-Australas J Anim Sci* **26**: 1144-1151.

Little E, Bork P & Doolittle R (1994) Tracing the spread of fibronectin type III domains in bacterial glycohydrolases. *Journal of Molecular Evolution* **39**: 631-643.

Lorenz P & Eck J (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* **3**: 510-516.

Lowe TM & Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955-964.

Lu J, Santo Domingo J & Shanks OC (2007) Identification of chicken-specific fecal microbial sequences using a metagenomic approach. *Water Res* **41**: 3561-3574.

Lytle BL, Volkman BF, Westler WM & Wu JH (2000) Secondary structure and calcium-induced folding of the Clostridium thermocellum dockerin domain determined by NMR spectroscopy. *Arch Biochem Biophys* **379**: 237-244.

Manichanh C, Rigottier-Gois L, Bonnaud E, *et al.* (2006) Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**: 205-211.

Markowitz VM, Ivanova N, Palaniappan K, *et al.* (2006) An experimental metagenome data management and analysis system. *Bioinformatics* **22**: e359-367.

Markowitz VM, Ivanova NN, Szeto E, *et al.* (2008) IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res* **36**: D534-538.

Markowitz VM, Chen IM, Chu K, *et al.* (2012) IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res* **40**: D123-129.

Maxam AMG, Walter (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* **74**: 560-564.

McKain N, Wallace RJ & Watt ND (1992) Selective isolation of bacteria with dipeptidyl aminopeptidase type I activity from the sheep rumen. *FEMS Microbiol Lett* **74**: 169-173.

Mesnage S, Fontaine T, Mignot T, Delepierre M, Mock M & Fouet A (2000) Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. *Embo j* **19**: 4473-4484.

Metzker ML (2009) Sequencing technologies—the next generation. *Nature Reviews Genetics* **11**: 31-46.

Meyer F, Paarmann D, D'Souza M, *et al.* (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Mitra S, Rupek P, Richter DC, Urich T, Gilbert JA, Meyer F, Wilke A & Huson DH (2011) Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. *BMC Bioinformatics* **12 Suppl 1**: S21.

Mizuguchi K, Dhanaraj V, Blundell TL & Murzin AG (1999) N-ethylmaleimide-sensitive fusion protein (NSF) and CDC48 confirmed as members of the double-psi beta-barrel aspartate decarboxylase/formate dehydrogenase family. *Structure* **7**: R215-216.

Molgaard A, Kauppinen S & Larsen S (2000) Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases. *Structure* **8**: 373-383.

Moon CD, Pacheco DM, Kelly WJ, Leahy SC, Li D, Kopecny J & Attwood GT (2008) Reclassification of Clostridium proteoclasticum as Butyrivibrio proteoclasticus comb. nov., a butyrate-producing ruminal bacterium. *Int J Syst Evol Microbiol* **58**: 2041-2045.

Mosbah A, Belaich A, Bornet O, Belaich JP, Henrissat B & Darbon H (2000) Solution structure of the module X2 1 of unknown function of the cellulosomal scaffolding protein CipC of Clostridium cellulolyticum. *J Mol Biol* **304**: 201-217.

Myers EW & Miller W (1988) Optimal alignments in linear space. *Comput Appl Biosci* **4**: 11-17.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M & Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**: 1344-1349.

Nava G, Bielke L, Callaway T & Castaneda M (2005) Probiotic alternatives to reduce gastrointestinal infections: the poultry experience. *Animal Health Research Reviews* **6**: 105-118.

Nordentoft S, Mølbak L, Bjerrum L, De Vylder J, Van Immerseel F & Pedersen K (2011) The influence of the cage system and colonisation of Salmonella Enteritidis on the microbial gut flora of laying hens studied by T-RFLP and 454 pyrosequencing. *BMC microbiology* **11**: 187.

Nuijens JH, van Berkel PH & Schanbacher FL (1996) Structure and biological actions of lactoferrin. *J Mammary Gland Biol Neoplasia* **1**: 285-295.

Oakley BB, Buhr RJ, Ritz CW, Kiepper BH, Berrang ME, Seal BS & Cox NA (2014) Successional changes in the chicken cecal microbiome during 42 days of growth are independent of organic acid feed additives. *BMC veterinary research* **10**: 282.

Oikonomou G, Machado VS, Santisteban C, Schukken YH & Bicalho RC (2012) Microbial diversity of bovine mastitic milk as described by pyrosequencing of metagenomic 16s rDNA. *PLoS One* **7**: e47671.

Orcutt B, Bailey B, Staudigel H, Tebo BM & Edwards KJ (2009) An interlaboratory comparison of 16S rRNA gene-based terminal restriction fragment length polymorphism and sequencing methods for assessing microbial diversity of seafloor basalts. *Environ Microbiol* **11**: 1728-1735.

Ott SJ, Musfeldt M, Wenderoth DF, Hampe J, Brant O, Folsch UR, Timmis KN & Schreiber S (2004) Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* **53**: 685-693.

Palackal N, Lyon CS, Zaidi S, *et al.* (2007) A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. *Appl Microbiol Biotechnol* **74**: 113-124.

Poinar HN, Schwarz C, Qi J, Shapiro B, MacPhee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP & Auch A (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**: 392-394.

Poole DM, Morag E, Lamed R, Bayer EA, Hazlewood GP & Gilbert HJ (1992) Identification of the cellulose-binding domain of the cellulosome subunit S1 from Clostridium thermocellum YS. *FEMS Microbiol Lett* **78**: 181-186.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J & Glöckner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188-7196.

Punta M, Coggill PC, Eberhardt RY, *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res* **40**: D290-301.

Qu A, Brulc JM, Wilson MK, Law BF, Theoret JR, Joens LA, Konkel ME, Angly F, Dinsdale EA & Edwards RA (2008) Comparative metagenomics reveals host specific metavirulomes and horizontal gene transfer elements in the chicken cecum microbiome. *PLoS One* **3**: e2945.

Ransom-Jones E, Jones DL, McCarthy AJ & McDonald JE (2012) The Fibrobacteres: an important phylum of cellulose-degrading bacteria. *Microb Ecol* **63**: 267-281.

Rees HC, Grant S, Jones B, Grant WD & Heaphy S (2003) Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles* **7**: 415-421.

Robert W (2012) The bacterial community composition of the bovine rumen detected using pyrosequencing of 16S rRNA genes. *Metagenomics* **2012**.

Rodrigue S, Materna AC, Timberlake SC, Blackburn MC, Malmstrom RR, Alm EJ & Chisholm SW (2010) Unlocking short read sequencing for metagenomics. *PLoS One* **5**: e11840.

Ross EM, Moate PJ, Bath CR, Davidson SE, Sawbridge TI, Guthridge KM, Cocks BG & Hayes BJ (2012) High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genet* **13**: 53.

Ross EM, Moate PJ, Bath CR, Davidson SE, Sawbridge TI, Guthridge KM, Cocks BG & Hayes BJ (2012) High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genet* **13**: 53.

Ross R, Mills S, Hill C, Fitzgerald G & Stanton C (2010) Specific metabolite production by gut microbiota as a basis for probiotic function. *International dairy journal* **20**: 269-276.

Sadet-Bourgeteau S, Martin C & Morgavi DP (2010) Bacterial diversity dynamics in rumen epithelium of wethers fed forage and mixed concentrate forage diets. *Vet Microbiol* **146**: 98-104.

Sanger F, Nicklen S & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**: 5463-5467.

Schloss PD & Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501-1506.

Schogor AL, Huws SA, Santos GT, Scollan ND, Hauck BD, Winters AL, Kim EJ & Petit HV (2014) Ruminal Prevotella spp. May Play an Important Role in the Conversion of Plant Lignans into Human Health Beneficial Antioxidants. *PLoS One* **9**: e87949.

Schuster SC (2007) Next-generation sequencing transforms today's biology. *Nature* **200**.

Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS & Huttenhower C (2013) Computational meta'omics for microbial community studies. *Molecular systems biology* **9**.

Sekelja M, Rud I, Knutsen S, Denstadli V, Westereng B, Næs T & Rudi K (2012) Abrupt temporal fluctuations in the chicken fecal microbiota are explained by its gastrointestinal origin. *Appl Environ Microbiol* **78**: 2941-2948.

Selengut JD, Haft DH, Davidsen T, Ganapathy A, Gwinn-Giglio M, Nelson WC, Richter AR & White O (2007) TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res* **35**: D260-264.

Shendure J & Ji H (2008) Next-generation DNA sequencing. *Nature biotechnology* **26**: 1135-1145.

Shoham Y, Lamed R & Bayer EA (1999) The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends Microbiol* **7**: 275-281.

Singh K, Shah T, Deshpande S, Jakhesara S, Koringa P, Rank D & Joshi C (2012) High through put 16S rRNA gene-based pyrosequencing analysis of the fecal microbiota of high FCR and low FCR broiler growers. *Molecular biology reports* **39**: 10595-10602.

Sjöling S, Stafford W & Cowan D (2007) Soil metagenomics: Exploring and exploiting the soil microbial gene pool.

SOUTO BdM, CARVALHO L, LOPIES N, CUNHA I, KRUGER R & QUIRINO B (2011) Identification of clones with Beta-Glucosidase and cellobiohydrolase activities from a goat rumen metagenomic library. p.^pp. In: BRAZILIAN BIOENERGY SCIENCE AND TECHNOLOGY CONFERENCE-BBEST, 1., 2011, Campos do Jordão, SP.[Anais...]. São Paulo: FAPESP, 2011.

Stappenbeck TS, Hooper LV & Gordon JI (2002) Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc Natl Acad Sci U S A* **99**: 15451-15455.

Stoddart D, Heron AJ, Mikhailova E, Maglia G & Bayley H (2009) Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proceedings of the National Academy of Sciences* **106**: 7702-7707.

Sugarbaker DJ, Richards WG, Gordon GJ, Dong L, De Rienzo A, Maulik G, Glickman JN, Chirieac LR, Hartman M-L & Taillon BE (2008) Transcriptome sequencing of malignant pleural mesothelioma tumors. *Proceedings of the National Academy of Sciences* **105**: 3521-3526.

Tajima K, Aminov RI, Nagamine T, Ogata K, Nakamura M, Matsui H & Benno Y (1999) Rumen bacterial diversity as determined by sequence analysis of 16S rDNA libraries. *FEMS Microbiology Ecology* **29**: 159-169.

Takase I, Ishino F, Wachi M, Kamata H, Doi M, Asoh S, Matsuzawa H, Ohta T & Matsuhashi M (1987) Genes encoding two lipoproteins in the leuS-dacA region of the Escherichia coli chromosome. *J Bacteriol* **169**: 5692-5699.

Tang Y, Underwood A, Gielbert A, Woodward MJ & Petrovska L (2014) Metaproteomics analysis reveals the adaptation process for the chicken gut microbiota. *Appl Environ Microbiol* **80**: 478-485.

Tews I, Perrakis A, Oppenheim A, Dauter Z, Wilson KS & Vorgias CE (1996) Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. *Nat Struct Biol* **3**: 638-648.

Tormo J, Lamed R, Chirino AJ, Morag E, Bayer EA, Shoham Y & Steitz TA (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *Embo j* **15**: 5739-5751.

Toyoda A, Iio W, Mitsumori M & Minato H (2009) Isolation and identification of cellulose-binding proteins from sheep rumen contents. *Appl Environ Microbiol* **75**: 1667-1673.

Trafalska E & Grzybowska K (2003) [Probiotics--an alternative for antibiotics?]. *Wiadomosci lekarskie (Warsaw, Poland: 1960)* **57**: 491-498.

Turnbaugh PJ, Backhed F, Fulton L & Gordon JI (2008) Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**: 213-223.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS & Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37-43.

van Gylswyk NO & vander Toorn JJTK (1986) Enumeration of Bacteroides succinogenes in the rumen of sheep fed maize-straw diets. *FEMS Microbiol Lett* **38**: 205-209.

Vandeplas S, Dubois Dauphin R, Beckers Y, Thonart P & Théwis A (2010) Salmonella in Chicken: Current and Developing Strategies To Reduce Contamination at Farm Level. *Journal of Food Protection* **73**: 774-785.

Videnska P, Faldynova M, Juricova H, Babak V, Sisak F, Havlickova H & Rychlik I (2013) Chicken faecal microbiota and disturbances induced by single or repeated therapy with tetracycline and streptomycin. *BMC veterinary research* **9**: 30.

Waldroup AL (1996) Contamination of raw poultry with pathogens. *World's Poultry Science Journal* **52**: 7-25.

Walker JE, Arizmendi JM, Dupuis A, Fearnley IM, Finel M, Medd SM, Pilkington SJ, Runswick MJ & Skehel JM (1992) Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from bovine heart mitochondria. Application of a novel strategy for sequencing proteins using the polymerase chain reaction. *J Mol Biol* **226**: 1051-1072.

Walker ND, McEwan NR & Wallace RJ (2003) Cloning and functional expression of dipeptidyl peptidase IV from the ruminal bacterium Prevotella albensis M384(T). *Microbiology* **149**: 2227-2234.

Wallace RJ & McKain N (1991) A survey of peptidase activity in rumen bacteria. *J Gen Microbiol* **137**: 2259-2264.

Wang G, Luo H, Meng K, Wang Y, Huang H, Shi P, Pan X, Yang P, Diao Q & Zhang H (2011) High genetic diversity and different distributions of glycosyl hydrolase family 10 and 11 xylanases in the goat rumen. *PLoS One* **6**: e16731.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J & Bähler J (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**: 1239-1243.

Wu S, Zhu Z, Fu L, Niu B & Li W (2011) WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC genomics* **12**: 444.

Xu GY, Ong E, Gilkes NR, Kilburn DG, Muhandiram DR, Harris-Brandts M, Carver JP, Kay LE & Harvey TS (1995) Solution structure of a cellulose-binding domain from Cellulomonas fimi by nuclear magnetic resonance spectroscopy. *Biochemistry* **34**: 6993-7009.

Yang Q, Ren S, Niu T, Guo Y, Qi S, Han X, Liu D & Pan F (2013) Distribution of antibiotic-resistant bacteria in chicken manure and manure-fertilized vegetables. *Environmental Science and Pollution Research* 1-11.

Yilmaz P, Iversen MH, Hankeln W, Kottmann R, Quast C & Glöckner FO (2012) Ecological structuring of bacterial and archaeal taxa in surface ocean waters. *FEMS Microbiology Ecology* **81**: 373-385.

Zhao M, Chen Y, Qu D & Qu H (2011) TSdb: a database of transporter substrates linking metabolic pathways and transporter systems on a genome scale via their shared substrates. *Sci China Life Sci* **54**: 60-64.

Zhou J, Bruns MA & Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62**: 316-322.

Zhou S (2014) Identification of the Microbial Composition and Cellulolytic Genes in Goat's Rumen Metagenome using Next-Generation Sequence Analysis. p.^pp. Plant and Animal Genome.

Zhu XY, Zhong T, Pandya Y & Joerger RD (2002) 16S rRNA-Based Analysis of Microbiota from the Cecum of Broiler Chickens. *Applied and Environmental Microbiology* **68**: 124-137.

Appendix

Appendix 1. List of minor phylum groups in each sample, whose portion were less than 1%.

Name	Rumen Liquid (%)	Rumen Solid (%)
Tenericutes	0.8623	0.0000
Lentisphaerae	0.4909	0.0000
Fibrobacteres	0.4776	0.0000
TM7	0.3184	0.5739
SR1	0.1990	0.1739
Elusimicrobia	0.0531	0.0522
Synergistetes	0.0531	0.0522
Verrucomicrobia	0.0398	0.1217
Bacteria_uc	0.0265	0.3826
Actinobacteria	0.0133	0.0174
BRM	0.0133	0.0174
Planctomycetes	0	0.2261
Chloroflexi	0	0.0174

Appendix 2. List of minor class groups in each sample, whose portion were less than 1%.

Name	Rumen Liquid (%)	Rumen Solid (%)
Alphaproteobacteria	0.9286	0
Mollicutes	0.8490	0.0000
Erysipelotrichi	0.8358	0.0000
Fibrobacteria	0.4511	0.0000
Lentisphaera_c	0.3184	0
TM7_c	0.3184	0.5391
SR1_c	0.1990	0.1739
Bacteroidetes_uc	0.1459	0.2783
AB185563_c	0.1459	0.1391
Firmicutes_uc	0.0796	0.2261
Bacilli	0.0663	0.0870
Proteobacteria_uc	0.0663	0.0522
Elusimicrobia_c	0.0531	0.0522
Synergistia	0.0531	0.0522
Opitutae	0.0398	0.0696
Bacteria_uc_c	0.0265	0.3826
Deltaproteobacteria	0.0265	0.1044
Spirochaetes_uc	0.0265	0.0348
Fibrobacteres_uc	0.0265	0.0348
Lentisphaerae_uc	0.0265	0
Acidobacteria_c	0.0265	0.0174
Cyanobacteria_uc	0.0133	0.0522
Actinobacteria_c	0.0133	0
Tenericutes_uc	0.0133	0.0174
BRM_c	0.0133	0.0174
Sphingobacteria	0.0133	0
AY289459_c	0.0133	0
EU844484_c	0.0133	0
Caldithrix_c	0.0133	0
Planctomycetacia	0	0.2261
Verrucomicrobiae	0	0.0348
Anaerolineae	0	0.0174
Verrucomicrobia_uc	0	0.0174
Flavobacteria	0	0.0174

Appendix 3. List of minor order groups in each sample, whose portion were less than 1%.

Name	Rumen Liquid (%)	Rumen Solid (%)
Bacteroidia_uc	0.9817	0
Spirochaetales	0.7562	0
Erysipelotrichales	0.7031	0
Clostridia_uc	0.7031	0
Alteromonadales	0.7031	0.0870
AM275436_o	0.5174	0.6435
Fibrobacterales	0.3847	0
Spirochaetes_c_uc	0.3184	0.0348
EU381732_o	0.3184	0.5044
AB185535_o	0.2919	0
AB185560_o	0.2653	0.4870
Acholeplasmatales	0.2521	0.7478
AY571491_o	0.2255	0.1913
SR1_o	0.1990	0.1739
Gammaproteobacteria_uc	0.1725	0.0696
Bacteroidetes_uc_o	0.1459	0.2783
Erysipelotrichi_uc	0.1327	0.1565
Terasakiella_o	0.1194	0.0870
Alphaproteobacteria_uc	0.1061	0.0870
AB185563_c_uc	0.1061	0
Firmicutes_uc_o	0.0796	0.2261
Mollicutes_uc	0.0663	0.4174
Fibrobacteria_uc	0.0663	0.3130
AB355089_c_uc	0.0663	0
Proteobacteria_uc_o	0.0663	0.0522
EU381783_o	0.0531	0.1044
Dethiosulfovibrio_o	0.0531	0.0348
Sphingomonadales	0.0531	0
Rhodospirillales	0.0531	0.0174
AB185563_o	0.0398	0.1217
Bacilli_uc	0.0398	0.0348
TG1_o	0.0398	0.0174
Xanthomonadales	0.0398	0.0174
Odyssella_o	0.0398	0
Bacteria_uc_o	0.0265	0.3826
Victivallales	0.0265	0.1044
Spirochaetes_uc_o	0.0265	0.5217
Fibrobacteres_uc_o	0.0265	0.0348
Opitutae_uc	0.0265	0
Lentisphaerae_uc_o	0.0265	0.2435
Bacillales	0.0265	0
EU335295_o	0.0265	0
EF445270_o	0.0133	0.1913
Myxococcales	0.0133	0.0870
Puniceicoccales	0.0133	0.0522
Cyanobacteria_uc_o	0.0133	0.0522
Deltaproteobacteria_uc	0.0133	0.0174
Enterobacteriales	0.0133	0.0174

Name	Rumen Liquid (%)	Rumen Solid (%)
BRM_o	0.0133	0.0174
Tenericutes_uc_o	0.0133	0.0174
Sphingobacteriales	0.0133	0.0174
AB300096_o	0.0133	0
Elusimicrobia_c_uc	0.0133	0
Rhodobacterales	0.0133	0
EU844484_o	0.0133	0
Caldithrix_c_uc	0.0133	0
Actinomycetales	0.0133	0
EU491744_o	0	0.0522
Aminobacterium_o	0	0.0174
AM162458_o	0	0.0174
EF999382_o	0	0.2957
Planctomycetales	0	0.2087
AB355089_c_uc	0	0.1217
Lactobacillales	0	0.0522
Anaeroplasmetales	0	0.0522
Rhizobiales	0	0.0348
Elusimicrobiales	0	0.0348
TM7_c_uc	0	0.0348
TM7_uc_o	0	0.0348
AB185563_c_uc	0	0.0174
Verrucomicrobiales	0	0.0174
Flavobacteriales	0	0.0174
Anaerolinaeles	0	0.0174
Opitutae_uc	0	0.0174
Verrucomicrobia_uc_o	0	0.0174
Planctomycetacia_uc	0	0.0174
Actinobacteria_c_uc	0	0.0174

Appendix 4. List of minor family groups in each sample, whose portion were less than 1%.

Name	Rumen Liquid (%)	Rumen Solid (%)
Bacteroidia_uc_f	0.9817	0
Clostridia_uc_f	0.7031	0
Selenomonas_f	0.6235	0.0522
Ruminococcaceae	0.5970	0
DQ673502_f	0.4643	0
Aeromonadales_uc	0.4643	0
Paraprevotella_f	0.4378	0
Spirochaetaceae	0.4113	0
Erysipelotrichaceae	0.3715	0
Fibrobacteraceae	0.3582	0
Spirochaetes_c_uc_f	0.3184	0.5217
EU381732_f	0.2919	0
Erysipelotrichales_uc	0.2919	0
Spirochaetales_uc	0.2786	0.8348
EU843262_f	0.2786	0
EF602759_f	0.2786	0.3652
AB185560_f	0.2388	0.4522
AB185535_f	0.1857	0
AM275436_o_uc	0.1857	0.1913
AB355089_o_uc	0.1857	0
EF445272_f	0.1857	0.0522
Gammaproteobacteria_uc_f	0.1725	0.0696
SR1_f	0.1725	0.0348
Acholeplasmataceae	0.1592	0.2783
Bacteroidetes_uc_f	0.1459	0.2783
AY571491_f	0.1459	0
DQ809955_f	0.1459	0
Erysipelotrichi_uc_f	0.1327	0.1565
AB185535_o_uc	0.1061	0
AF371913_f	0.1061	0.0870
Alphaproteobacteria_uc_f	0.1061	0.0870
AB185563_c_uc_f	0.1061	0.0174
EU460065_f	0.0929	0
Acholeplasmatales_uc	0.0929	0.4696
EU462377_f	0.0796	0.5391
Firmicutes_uc_f	0.0796	0
Alteromonadales_uc	0.0796	0.0348
AY571491_o_uc	0.0796	0.1913
EU381764_f	0.0663	0
EU381456_f	0.0663	0.9913
4P000122_f	0.0663	0.5739
Mollicutes_uc_f	0.0663	0.4174
Fibrobacteria_uc_f	0.0663	0
AB355089_c_uc_f	0.0663	0.1739
EF436386_f	0.0663	0.0522
Proteobacteria_uc_f	0.0663	0.0522

Name	Rumen Liquid (%)	Rumen Solid (%)
AM275436_f	0.0531	0
EU381813_f	0.0531	0.1217
Terasakiella_o_uc	0.0531	0.0348
AF371945_f	0.0531	0.0174
Sphingomonadaceae	0.0531	0.0174
4P000161_f	0.0398	0.5044
4P000310_f	0.0398	0.0870
EU465874_f	0.0398	0.0522
FJ367060_f	0.0398	0.0348
Bacilli_uc_f	0.0398	0.0348
TG1_f	0.0398	0.0174
Xanthomonadaceae	0.0398	0.0174
Odyssella_o_uc	0.0398	0
Rhodospirillaceae	0.0398	0
Fibrobacterales_uc	0.0265	0.4000
Bacteria_uc_f	0.0265	0.3826
EU381859_f	0.0265	0
Victivallaceae	0.0265	0.1044
Mogibacterium_f	0.0265	0.0870
EU381783_f	0.0265	0.0870
AB185560_o_uc	0.0265	0
Spirochaetes_uc_f	0.0265	0.0348
Fibrobacteres_uc_f	0.0265	0.3130
Pyramidobacter_f	0.0265	0.0348
Opitutae_uc_f	0.0265	0.0174
EU381783_o_uc	0.0265	0.0174
EU381732_o_uc	0.0265	0
Lentisphaerae_uc_f	0.0265	0.2435
EU471633_f	0.0265	0
EU335295_f	0.0265	0
AB185519_f	0.0265	0
Dethiosulfovibrio_o_uc	0.0265	0
EF445270_f	0.0133	0
AB185563_f	0.0133	0
Cyanobacteria_uc_f	0.0133	0
AB185724_f	0.0133	0.0348
Bacteroidaceae	0.0133	0.0174
Deltaproteobacteria_uc_f	0.0133	0.0174
Enterobacteriaceae	0.0133	0.0174
AB185570_f	0.0133	0.0174
Tenericutes_uc_f	0.0133	0.0174
Rhodospirillales_uc	0.0133	0.0174
AB034150_f	0.0133	0
AB034027_f	0.0133	0
AB300096_o_uc	0.0133	0
Elusimicrobia_c_uc_f	0.0133	0
EU844484_f	0.0133	0

Name	Rumen Liquid (%)	Rumen Solid (%)
BRM_f	0.0133	0
Bacillales_uc	0.0133	0
Sphingobacteriaceae	0.0133	0
Corynebacteriaceae	0.0133	0
Paenibacillaceae	0.0133	0
Caldithrix_c_uc_f	0.0133	0
Rhodobacterales_uc	0.0133	0
Veillonellaceae	0	0.8870
EU381732_f	0	0.5044
EU843262_f	0	0.4348
Erysipelotrichales_uc	0	0.3304
Aeromonadales_uc	0	0.2609
AM275436_o_uc	0	0.2609
Firmicutes_uc_f	0	0.2261
EF999382_o_uc	0	0.2087
DQ809955_f	0	0.1739
AB185535_o_uc	0	0.1739
Planctomycetaceae	0	0.1565
EF445270_f	0	0.1391
EU381859_f	0	0.1217
AB355089_c_uc_f	0	0.1217
EU471817_f	0	0.0870
AB185563_f	0	0.0870
Puniceicoccales_uc	0	0.0522
Planctomycetales_uc	0	0.0522
DQ809526_f	0	0.0522
Cyanobacteria_uc_f	0	0.0522
Myxococcales_uc	0	0.0522
EU491744_o_uc	0	0.0522
Anaeroplasmetales_uc	0	0.0522
AB185560_o_uc	0	0.0348
AB185563_o_uc	0	0.0348
EU844239_f	0	0.0348
Fibrobacteres_uc_f	0	0.0348
TM7_c_uc_f	0	0.0348
DQ326408_f	0	0.0348
Rhizobiales_uc	0	0.0348
TM7_uc_f	0	0.0348
Elusimicrobiales_uc	0	0.0348
Streptococcaceae	0	0.0348
EF436357_f	0	0.0174
Anaerolinaceae	0	0.0174
Leuconostocaceae	0	0.0174
Verrucomicrobia_uc_f	0	0.0174
SR1_o_uc	0	0.0174
Planctomycetacia_uc_f	0	0.0174
Actinobacteria_c_uc_f	0	0.0174

Name	Rumen Liquid (%)	Rumen Solid (%)
BRM_o_uc	0	0.0174
AB186799_f	0	0.0174
Flavobacteriales_uc	0	0.0174
EF445140_f	0	0.0174
AM162458_o_uc	0	0.0174

Appendix 5. List of minor genus groups in each sample, whose portion were less than 1%.

Name	Rumen Liquid (%)	Rumen Solid (%)
Bacteroidia_uc_g	0.9817	0
AB270013_g	0.8623	0.8870
Succiniclasticum	0.8490	0.7304
Succinivibrionaceae_uc	0.7562	0.4174
Clostridia_uc_g	0.7031	0
Ruminobacter	0.5970	0.3652
Selenomonas	0.5704	0.0522
EU844736_g	0.5439	0.7130
Aeromonadales_uc_g	0.4643	0.2609
EU845084_f_uc	0.4378	0.5044
Veillonellaceae_uc	0.4378	0.1044
Butyrivibrio	0.3317	0
Fibrobacter	0.3184	0
Spirochaetes_c_uc_g	0.3184	0.5217
AB355089_f_uc	0.2919	0.3304
Erysipelotrichales_uc_g	0.2919	0.3304
Ruminococcaceae_uc	0.2919	0.7304
AB185751_f_uc	0.2786	0.0870
Spirochaetales_uc_g	0.2786	0.8348
Paraprevotella_f_uc	0.2653	0
DQ673502_f_uc	0.2521	0.8000
AB185560_g	0.2388	0.3826
Erysipelotrichaceae_uc	0.2388	0.3478
Anaerovibrio	0.2255	0.0522
Treponema_g1	0.2255	0.9565
DQ673502_g	0.2123	0.6957
Pseudobutyrivibrio	0.2123	0.5391
EU843169_g	0.1990	0.1913
AB355089_o_uc_g	0.1857	0.1739
AM275436_o_uc_g	0.1857	0.2609
Spirochaetaceae_uc	0.1857	0.5913
EF445272_f_uc	0.1725	0
EU381732_g	0.1725	0.3478
EU843262_g	0.1725	0.2261
Gammaproteobacteria_uc_g	0.1725	0.0696
EU622705_g	0.1592	0.1391
Bacteroidetes_uc_g	0.1459	0.2783
AB009195_g	0.1327	0
Erysipelotrichi_uc_g	0.1327	0.1565
Ruminococcus	0.1327	0.2087
SR1_g	0.1327	0.0348
Succinimonas	0.1327	0.0174
AB185535_o_uc_g	0.1061	0.1739
AB185563_c_uc_g	0.1061	0.0174
Acholeplasma	0.1061	0.1391
Alphaproteobacteria_uc_g	0.1061	0.0870
EF602759_f_uc	0.1061	0.1739
EU622729_g	0.1061	0.0696

Name	Rumen Liquid (%)	Rumen Solid (%)
EU843262_f_uc	0.1061	0.2087
EU844703_g	0.1061	0.3478
Acholeplasmatales_uc_g	0.0929	0.4696
EF445280_g	0.0929	0.1391
EU843283_g	0.0929	0.0696
EU844830_g	0.0929	0.7478
4P000075_g	0.0796	0
Alteromonadales_uc_g	0.0796	0.0348
AY571491_o_uc_g	0.0796	0
DQ809955_g	0.0796	0.1217
EF404684_g	0.0796	0.2435
EU259443_g	0.0796	0.1217
Firmicutes_uc_g	0.0796	0.2261
Lachnobacterium	0.0796	0
Schwartzia	0.0796	0
4P000122_g	0.0663	0.4696
AB269949_g	0.0663	0
AB355089_c_uc_g	0.0663	0.1217
AF371913_g	0.0663	0.0348
EU460065_f_uc	0.0663	0.8174
Fibrobacteria_uc_g	0.0663	0.3130
Mollicutes_uc_g	0.0663	0.4174
Proteobacteria_uc_g	0.0663	0.0522
Roseburia	0.0663	0.1913
Acholeplasmataceae_uc	0.0531	0.1391
AF371945_g	0.0531	0
DQ809955_f_uc	0.0531	0.0522
EF436386_f_uc	0.0531	0.0522
EF686516_g	0.0531	0.0870
EU381456_g	0.0531	0.8348
EU381640_g	0.0531	0
EU844535_g	0.0531	0.1217
EU844657_g	0.0531	0.0696
EU844681_g	0.0531	0.0174
Eubacterium_g7	0.0531	0.3304
Selenomonas_f_uc	0.0531	0
Sphingomonas	0.0531	0.0174
Terasakiella_o_uc_g	0.0531	0.0348
4P000161_f_uc	0.0398	0.1217
4P000310_f_uc	0.0398	0.0870
AB185535_f_uc	0.0398	0.3130
AF371913_f_uc	0.0398	0.0522
AY571491_f_uc	0.0398	0.0348
Bacilli_uc_g	0.0398	0.0348
EF403870_g	0.0398	0.3130
EU381764_f_uc	0.0398	0.8522
EU462208_g	0.0398	0.0348
EU462377_f_uc	0.0398	0.1565
EU462623_g	0.0398	0.0696

Name	Rumen Liquid (%)	Rumen Solid (%)
EU777390_g	0.0398	0.2261
Fibrobacteraceae_uc	0.0398	0.4000
Odyssella_o_uc_g	0.0398	0
SR1_f_uc	0.0398	0
AB034061_g	0.0265	0
AB185519_f_uc	0.0265	0
AB185548_g	0.0265	0
AB185560_o_uc_g	0.0265	0.0348
AB185756_g	0.0265	0.0174
Anaerosporobacter	0.0265	0
Bacteria_uc_g	0.0265	0.3826
Dethiosulfovibrio_o_uc_g	0.0265	0
EF445265_g	0.0265	0
EU335295_f_uc	0.0265	0
EU335363_g	0.0265	0
EU381732_f_uc	0.0265	0.0174
EU381732_o_uc_g	0.0265	0
EU381764_g	0.0265	0.3478
EU381783_g	0.0265	0.0870
EU381783_o_uc_g	0.0265	0.0174
EU381813_f_uc	0.0265	0.0696
EU381813_g	0.0265	0.0522
EU381859_g	0.0265	0.1044
EU382030_g	0.0265	0.0348
EU382060_g	0.0265	0.0174
EU462377_g	0.0265	0.3826
EU471633_f_uc	0.0265	0
EU842564_g	0.0265	0.0696
EU843437_g	0.0265	0.0348
EU845084_g	0.0265	0.0174
Fibrobacterales_uc_g	0.0265	0.4000
Fibrobacteres_uc_g	0.0265	0.0348
FJ172836_g	0.0265	0.0174
FJ367060_f_uc	0.0265	0.0348
Lentisphaerae_uc_g	0.0265	0
Mogibacterium_f_uc	0.0265	0.0348
Opitutae_uc_g	0.0265	0.0174
Spirochaetes_uc_g	0.0265	0.0348
Xanthomonadaceae_uc	0.0265	0.0174
4P000387_g	0.0133	0
AB034027_f_uc	0.0133	
AB034150_g	0.0133	
AB185535_g	0.0133	0.2261
AB185570_g	0.0133	0.0174
AB185578_g	0.0133	0.0348
AB185724_f_uc	0.0133	0.0174
AB300096_o_uc_g	0.0133	0
AJ428412_g	0.0133	0.0174
AM275436_f_uc	0.0133	0.1044

Name	Rumen Liquid (%)	Rumen Solid (%)
AM275436_g	0.0133	0
AM696934_g	0.0133	0
AY571491_g	0.0133	0.0174
AY976335_g	0.0133	0
Bacillales_uc_g	0.0133	0
Bacteroides	0.0133	0
BRM_f_uc	0.0133	0
Caldithrix_c_uc_g	0.0133	0
Corynebacterium	0.0133	0
Cyanobacteria_uc_g	0.0133	0.0522
Deltaproteobacteria_uc_g	0.0133	0.0174
DQ394602_g	0.0133	0
DQ394617_g	0.0133	0
DQ673567_g	0.0133	0
DQ905714_g	0.0133	0
EF404362_g	0.0133	0.0348
EF404633_g	0.0133	0
EF445268_g	0.0133	0.0348
EF445270_g	0.0133	0.1217
Elusimicrobia_c_uc_g	0.0133	0.0348
Erwinia	0.0133	0
EU259446_g	0.0133	0.9739
EU259455_g	0.0133	0
EU381456_f_uc	0.0133	0.1565
EU381760_g	0.0133	0.0348
EU381810_g	0.0133	0
EU453178_g	0.0133	0
EU459572_g	0.0133	0
EU842803_g	0.0133	0
EU844484_f_uc	0.0133	0
Paenibacillaceae_uc	0.0133	0
Pedobacter	0.0133	0
Rhodobacterales_uc_g	0.0133	0
Rhodospirillales_uc_g	0.0133	0.0174
Skermanella	0.0133	0
Stenotrophomonas	0.0133	0.0174
Tenericutes_uc_g	0.0133	0.0174
TG1_f_uc	0.0133	0
4P000122_f_uc	0	0.1044
4P000161_g	0	0.3826
AB034001_g	0	0.0174
AB034016_g	0	0.0174
AB185516_g	0	0.1739
AB185524_g	0	0.0174
AB185560_f_uc	0	0.0696
AB185563_f_uc	0	0.0522
AB185563_o_uc_g	0	0.0348
AB185724_g	0	0.0174
AB186799_f_uc	0	0.0174

Name	Rumen Liquid (%)	Rumen Solid (%)
AB270056_g	0	0.0174
Actinobacteria_c_uc_g	0	0.0174
AF018497_g	0	0.1327
AF371945_f_uc	0	0.0174
AF499908_g	0	0.0348
AM162458_o_uc_g	0	0.0174
Anaeroplasmetales_uc_g	0	0.0522
AY212760_g	0	0.0348
AY442825_g	0	0.0174
AY854276_g	0	0.0174
Bacteroidaceae_uc	0	0.0174
BRM_o_uc_g	0	0.0174
Clostridium_g11	0	0.0174
Coprococcus	0	0.0348
DQ326408_g	0	0.0348
DQ394643_g	0	0.0348
DQ532165_g	0	0.0174
DQ799133_g	0	0.0174
DQ808472_g	0	0.0174
DQ809526_f_uc	0	0.0522
EF436334_g	0	0.0174
EF436357_f_uc	0	0.0174
EF436358_g	0	0.0174
EF445140_f_uc	0	0.0174
EF445270_f_uc	0	0.0174
EF445270_o_uc_g	0	0.0522
EF445281_g	0	0.0174
EF999382_o_uc_g	0	0.2087
Enterobacter	0	0.0174
EU259378_g	0	0.0174
EU381489_g	0	0.0348
EU381491_g	0	0.0348
EU381832_g	0	0.0174
EU381859_f_uc	0	0.0174
EU459524_g	0	0.0696
EU459537_g	0	0.0174
EU463782_g	0	0.0174
EU465874_f_uc	0	0.0174
EU471634_g	0	0.0174
EU471817_f_uc	0	0.0174
EU491744_o_uc_g	0	0.0522
EU622778_g	0	0.2609
EU777575_g	0	0.0348
EU842575_g	0	0.0174
EU843299_g	0	0.0174
EU844239_f_uc	0	0.0174
EU844239_g	0	0.0174
Eubacterium_g4	0	0.0174
Flavobacteriales_uc_g	0	0.0174

Name	Rumen Liquid (%)	Rumen Solid (%)
Lactococcus	0	0.0174
Lentisphaera_c_uc_g	0	0.2435
Mogibacterium	0	0.0522
Myxococcales_uc_g	0	0.0522
Planctomycetaceae_uc	0	0.1044
Planctomycetacia_uc_g	0	0.0174
Planctomycetales_uc_g	0	0.0522
Puniceicoccales_uc_g	0	0.0522
Rhizobiales_uc_g	0	0.0348
SR1_o_uc_g	0	0.0174
TM7_c_uc_g	0	0.0348
TM7_uc_g	0	0.0348
Verrucomicrobia_uc_g	0	0.0174
Victivallaceae_uc	0	0.0696
Weissella	0	0.0174

Appendix 6. List of minor (<1%) genus groups in each sample

Name	1 day	35 days
AB239481_f_uc	0	0.0452
AB606262_g	0	0.0226
AB626898_g	0	0.0678
AB626922_g	0	0.0226
Acinetobacter	0	0.0226
Aerococcus	0	0.0452
AF018558_g	0	0.0452
AF544207_f_uc	0	0.0226
AJ279038_g	0	0.0226
Alkaliphilus	0	0.0226
AM275436_o_uc_g	0	0.0678
AM277340_g	0	0.0678
AM406061_g	0	0.0226
Anaerofilum	0	0.0226
Arcobacter	0	0.0226
Bacillus	0	0.0226
Bacteroides	0	0.0452
Blautia	0.01424	0.70056
Brevundimonas	0	0.0226
Butyricicoccus	0	0.13559
Carnobacteriaceae_uc	0.00712	0
Cellulosilyticum	0	0.0904
Clostridia_uc_g	0.00712	0.0226
Clostridiaceae_uc	0.03559	0
Clostridiales_uc_g	0.00712	0.29379
Clostridium	0	0.54237
Clostridium_g16	0	0.0226
Clostridium_g23	0	0.0678
Clostridium_g6	0.37728	0.0904
Clostridium_g7	0	0.0226
Clostridium_g9	0	0.31638
Comamonas	0	0.0226
Coprobacillus	0.02136	0.0226
Corynebacterium	0	0.31638
Cronobacter	0.00712	0
Dorea	0	0.27119
DQ071456_g	0	0.13559
DQ071484_g	0	0.22599
DQ456434_g	0	0.0226
EF400272_g	0	0.11299
EF406589_g	0	0.0226
EF445272_f_uc	0	0.0226
EF604822_g	0	0.0226
Enterobacter	0.01424	0
Enterobacteriaceae_uc	0.02136	0

Name	1 day	35 days
Enterococcaceae_uc	0.02136	0
Enterococcus	0	0.22599
Epulopiscium	0.01424	0
Escherichia	0	0.29379
EU381725_g	0	0.29379
Eubacterium_g5	0	0.0678
Facklamia	0	0.0678
Faecalibacterium	0	0.90395
Festuca	0.00712	0
FJ966226_g	0	0.24859
Gallicola	0	0.18079
Gammaproteobacteria_uc_g	0.00712	0
GQ175418_g	0	0.0904
GQ451199_g	0	0.0226
GQ897562_g	0	0.0226
GQ897654_g	0	0.0226
GQ898349_g	0	0.0226
GU324404_g	0	0.0226
Hedyosmum	0.03559	0
HM124144_g	0	0.0226
HM124151_f_uc	0	0.0452
HM124260_g	0	0.0226
HQ452860_g	0	0.18079
Jeotgalicoccus	0	0.0226
Klebsiella	0.00712	0
Kurthia	0	0.13559
Lachnospiraceae_uc	0	0.47458
Lactobacillaceae_uc	0.00712	0.15819
Lactobacillales_uc_g	0	0.0226
Lactonifactor	0	0.29379
Lysinibacillus	0	0.0904
Mogibacterium_f_uc	0	0.0226
Nicotiana	0.01424	0
Oscillibacter	0	0.0904
Pelomonas	0.00712	0
Peptostreptococcaceae_uc	0	0.0678
Phaseolus	0.0783	0
Planococcaceae_uc	0	0.0226
Poaceae_uc	0.00712	0
Proteobacteria_uc_g	0.00712	0
Pseudoflavonifractor	0	0.18079
Ruminococcaceae_uc	0	0.47458
Ruminococcus_g2	0	0.0226
Ruminococcus_g3	0	0.11299
Ruminococcus_g4	0	0.31638
Secale	0.00712	0
Soonwooa	0	0.0226

Name	1 day	35 days
Sphingobacterium	0	0.0226
Staphylococcus	0	0.22599
Subdoligranulum	0	0.20339
Syntrophococcus	0	0.11299
Thermohalobacter_f_uc	0	0.0452
Trichococcus	0	0.0904
Triticum	0.00712	0
Wautersiella	0	0.0226
Weissella	0.00712	0.0226

요약(국문초록)

차세대 시퀀싱을 이용한 가축들의 메타제놈에 대한 연구

임 수 연

농생명공학부 동물생명공학전공

서울대학교 대학원 농업생명과학대학

본 본 학위논문에서는 가축 내 미생물 군집구조에 대한 메타제놈과 그 특성을 다양한 차세대 시퀀싱과 생물정보학 기법을 이용하여 분석하였다. 구체적으로는 소/염소의 반추위 및 닭의 분변 시료내의 미생물 군집구조를 분석한 후, 각각에 대하여 계통분류학적 특성 및 유전자를 확인하고자 하였다. 반추위 미생물의 제놈 분석, 유전자 역할 및 응용 가능성의 확인은 반추위 생태를 이해하는 데 도움이 된다. 닭의 분변은 성장과정 중 가금류의 오염과 위생과 밀접한 관련이 있다. 염소의 경우에는 다양한 섬유질 소화효소를 생산해 거친 풀도 쉽게 소화할 수 있기 때문에 반추위 내 미생물 군집은 특별한 섬유질 소화효소 활성을 보유할 것으로 예상된다.

제 1 장에서는 차세대 염기 서열 분석법과 가축에 적용한 메타제놈 연구에 관한 배경지식 및 기술동향을 요약하였다.

제 2 장에서는 차세대 시퀀싱 방법을 이용하여 소에서의 반추위 액과 반추위 내의 건초에서 분리한 미생물 군집구조를 분석하였다. 반추위 포유류에서 흔히 발견되는 미생물 군집이 밝혀지었고 급여를 했을 때의 건초에서 특이적으로 존재하는 미생물이 있음을 알아낼 수 있었다. 구체적으로는 **Bacteroidetes, Firmicutes** 그리고 **Fibrobacter** 가 반추위액보다 건초에서 보다 높은 비율로 분포함을 확인하였고 이 결과를 바탕으로 위 군집이 식물 섬유질을 분해하는데 중요한 역할을 함을 유추할 수 있었다. 또한 반추위 대사에 중요한 역할을 하는 ***Proteobacteria*** 와 당분대사에 관여하는 ***Mollicutes,*** 단백질 분해에 관여하는 ***Prevotellacea*** *등을 확인하였다.*

제 3 장에서는 차세대 시퀀싱 방법을 이용하여 35 일간의 생장 기간 동안 조류인 닭의 분변에서의 미생물을 밝혔다. 그 결과 5 주동안 생장기간 동안 닭 분변의 미생물 군집구조는 속 단위에서 30 개에서 87 개로 증가함을 확인할 수 있었다. 또한 이러한 다양성에도 불구하고, 80% 가 넘는 군집구조가 각각의 생장 주기에서 서로 다른 소수의 미생물 종으로만 이루어져 있음을 알 수 있었다. 상기 결과로 보아 닭의 분변의 미생물 군집은 제한된 분류 군으로

구성되어있지만 짧은 성장기간 동안 급격히 변화하였음을 확인할 수 있다.

제 4 장에서는 차세대 시퀀싱방법을 이용하여 반추위 동물인 염소의 장내 미생물 군집구조에서 셀룰로오즈 분해에 관여되는 특정 단백질 도메인들에 대하여 연구하였다. 염소 반추위 미생물 유전체들은 **shotgun sequencing** 을 통하여 구축되었고 **METAIDBA** 를 이용하여 선별되었다. 그 결과 미생물 군집은 총 **439** 개의 속으로 이루어지며, 그 중 ***Prevotella*** 와 ***Butyrivibrio*** 이 우점종으로 확인되었다. 그와 동시에 셀룰레이즈와 관련된 **201** 개의 시퀀스를 **NCBI** 데이터 베이스에서 검색하여 그 중 **28** 개의 단백질 도메인들을 **HMMER** 을 통하여 확인할 수 있었다. 염소 내 미생물 군집구조에서 셀룰레이즈 관련 단백질들을 관련 지어본 결과 **lipase GDSL, celluase** 그리고 **Glyco hydro 10** 가 주요 단백질 도메인임을 알 수 있었다. 이러한 단백질들간의 상관관계 또한 찾아본 결과 밀집한 양의 관계를 띄고 있음을 알 수 있었고, 이는 반추위 미생물이 오랜 시간 적응되어 특정 기능을 수행하는 군집들이 우세하게 존재한다는 것을 찾아내었다.

최근 연구결과들에서는 장내 미생물 환경이 장내 미생물의 군집구조 분포와 비만 및 성장 상태와 같은 숙주의 대사 특성과 밀접한 관계가 있음이 밝혀지고 있다. 따라서, 소, 닭, 염소와 같이

다양한 주요 가축내에서의 장내 미생물 군집을 메타제놈으로 접근하여 분석한 본 연구는 축산업에서의 생산성과 지속가능성에 향상 측면에서 활용이 가능하다.

주요어: 차세대 시퀀싱, 메타제놈, 반추동물, 미생물 군집구조
학번: 2010-30316